ETF3231/5231 Individual Assignment 3

Syed Jazil Hussain 28900766

```r
# Read in and tidy up your data
# Make sure you select the column with your student ID

# First three rows contain metadata, read them in separately
meta <- read_csv("Undergrad_Data.csv", col_names = TRUE, n_max = 3)
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_character()
## )
## i Use `spec()` for the full column specifications.
```

```r
# meta

# The data follows after the third row, we skip the metadata and read the data.
# Note: Skipping the first row skips the column names, we add them back from the
#       metadata.
dat <- read_csv("Undergrad_Data.csv",
                # use column names from the metadata
                col_names = colnames(meta),
                # skip 4 rows as we also skip column names, specified above
                skip = 4,
                # The automatic column types correctly guess all columns but the
                # date, we specify the date format manually here to correctly
                # get dates.
                col_types = cols("Student ID" = col_date("%b-%y")))

my_series <- dat %>%
  # feel free to rename your series appropriately
  rename(Month = "Student ID", y ="28900766") %>%
  select(Month, y) %>%
  mutate(Month=yearmonth(Month)) %>%
  as_tsibble(index = Month)
```

For the tasks that follow you will be modeling the **original data without any transformations** performed (even if you previously deemed it necessary).

Question 1

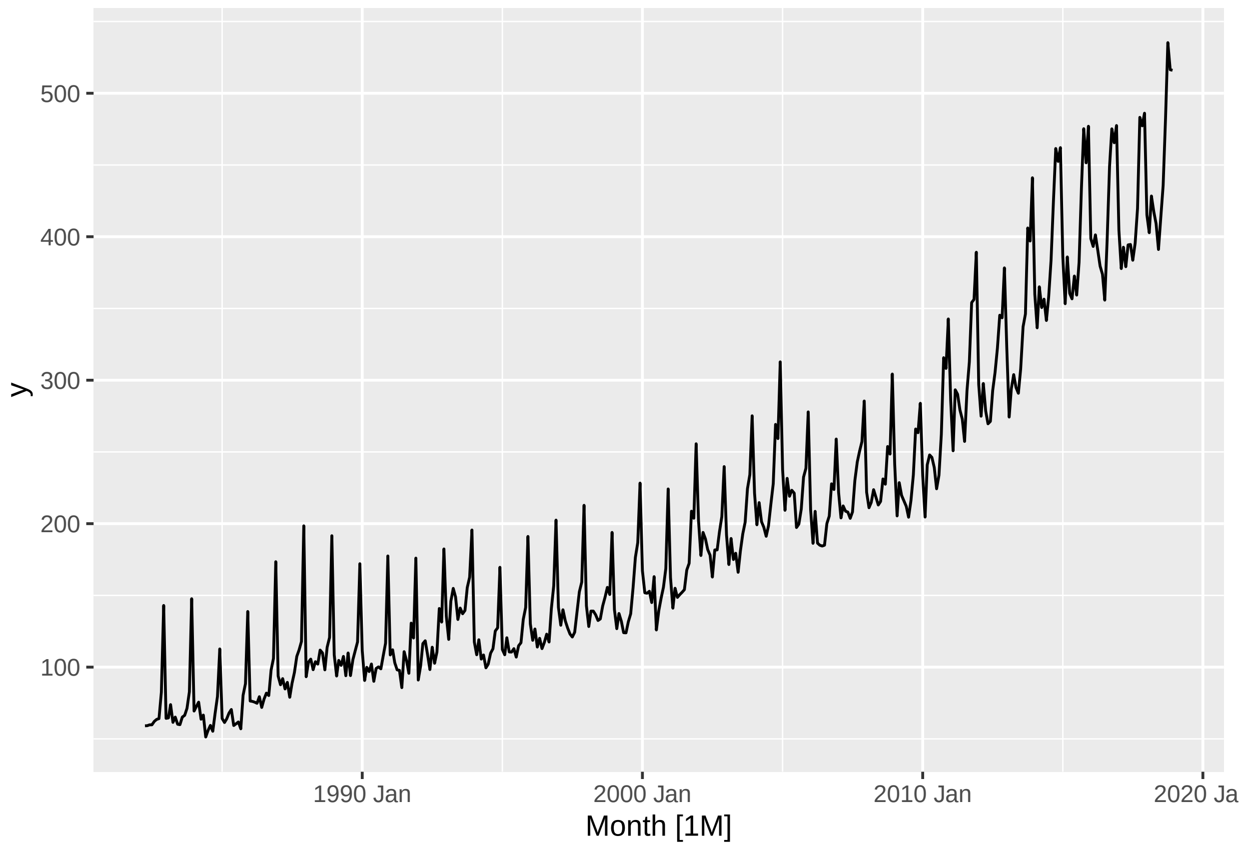
Plot your time series. By observing the plot and describing its components select an ETS model you think is appropriate for forecasting. Make sure you justify your choice (no more than 150 words). (8 marks)

The data set 'my__series' shows a positive trend throughout the years. The time series plot also shows multiplicative seasonality yearly where the components multiply together to make the time series, an increasing trend, the amplitude of seasonal activity increase. (Seasonal series is influenced by seasonal factors i.e year.) The seasonal effect also changes as time increases. There is no cyclic pattern. The crest represents the peak of the turnover which is in November initially but shifted to November in the latter years. The troughs represented the lowest turnover in the year which was in February throughout the years.

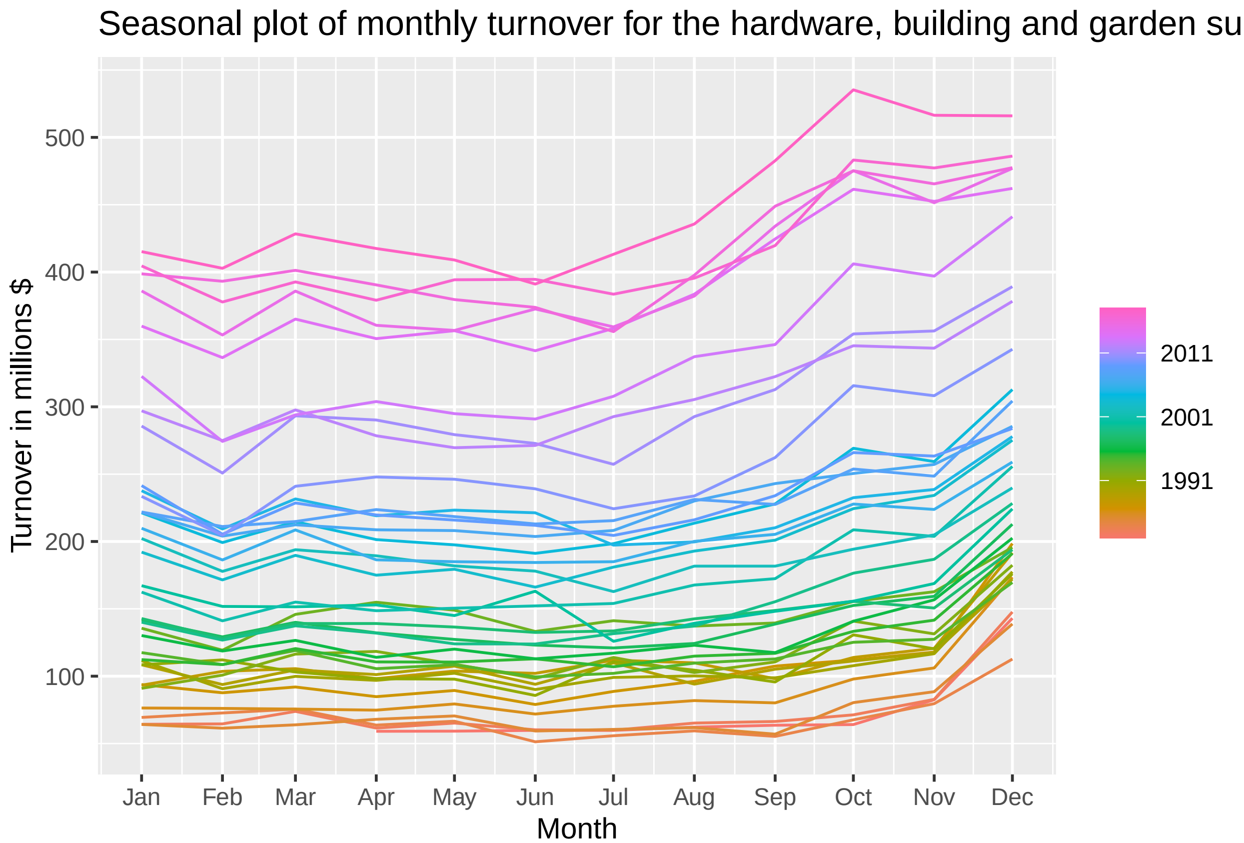
The ETS model that is appropriate for forecasting is a [Ad M] model (Holt-Winters' damped method) because the trend component for the time series is an additive decomposition while the seasonal component is multiplicative. However, the error cannot be determined from the given time series plot.

```
autoplot(my_series)
```

```
## Plot variable not specified, automatically selected `.vars = y`
```




```
my_series %>%
  gg_season(y)+ labs(y="Turnover in millions $", title= "Seasonal plot of monthly turnover for the hardw
```


Seasonal plot of monthly turnover for the hardware, building and garden su
500
400
300
200
100
Turnover in millions $
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
2011
2001
1991

Question 2

Estimate the ETS model you described in Question 1 and show the estimated model output. Describe and comment on the estimated parameters and components. Include any plots you see necessary (no more than 150 words). (14 marks)

Holt-winters' damped method:

Alpha smoothing parameter: Is 0.4996. Which can be said it's relatively large, therefore, the forecasted past value of the level is not dependent but rather depends on current observation at time t.

Beta smoothing parameter: Is 0.0028, which is very close to 0, indicating that the trend does not adjust to the level. Therefore, the trend component is not affected by what has changed in the level.

Gamma smoothing parameter: Is 0.2595, which is greater than 0, indicating that seasonal variation does change with time.

Phi smoothing parameter: Is 0.9388, which is close to 1, thus, shows a relatively strong dampening effect. Even though the trend is not linear, the strong dampening effect will slowly bring the forecasted to linear because phi=1.

Components:

Level and slope components follow the trend in the data fairly closely. The seasonal component displays a variation in seasonality which decreases over time which shows that it is multiplicative. The robustness weights is based on the remainder.

```r
fit<- my_series %>%
  model(

    hwdamped=ETS(y~error("M")+trend("Ad")+season("M")),
  )
fit %>% glance()
```

```
## # A tibble: 1 x 9
##   .model    sigma2 log_lik   AIC  AICc   BIC   MSE  AMSE    MAE
##   <chr>      <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  <dbl>
## 1 hwdamped 0.00396  -2377. 4790. 4792. 4863.  121.  166. 0.0464
```

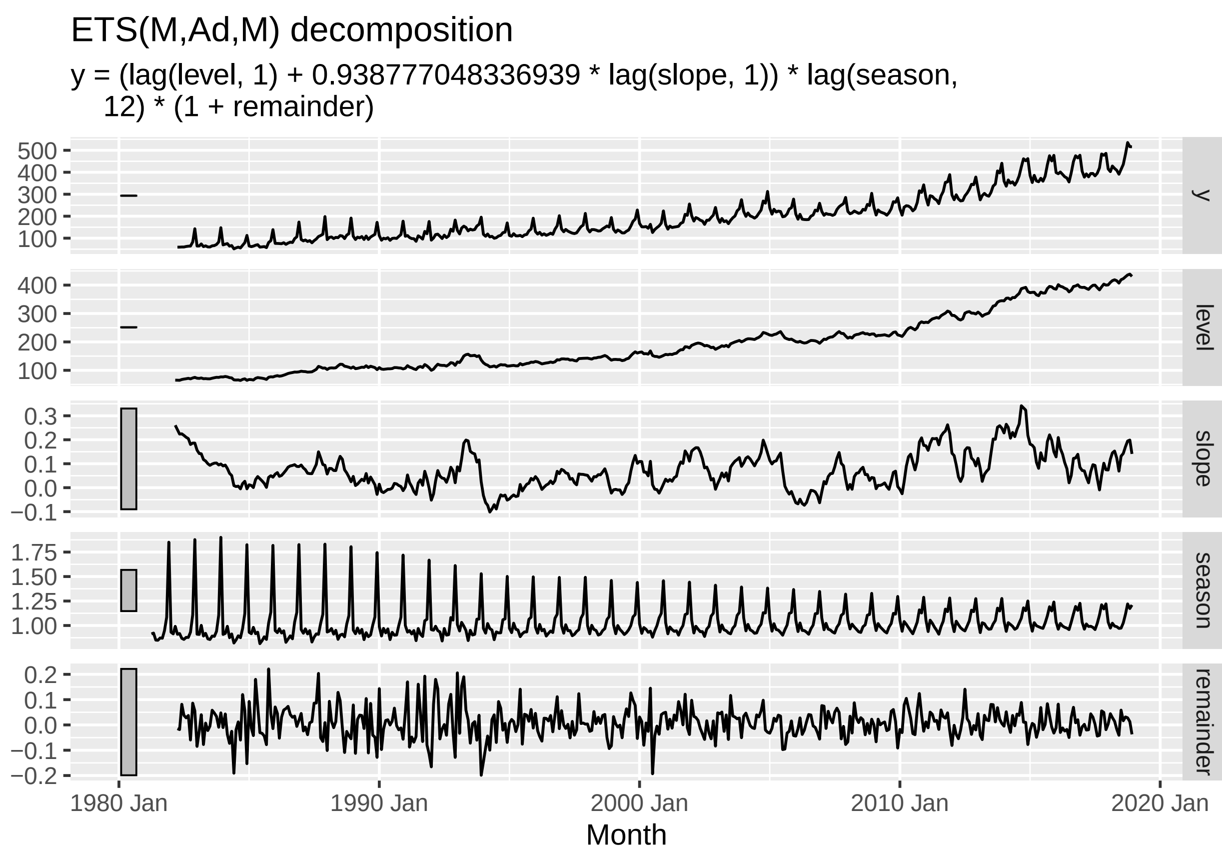
```r
fit %>%
  report(fit)
```

```
## Series: y
## Model: ETS(M,Ad,M)
##   Smoothing parameters:
##     alpha = 0.4996205
##     beta  = 0.002797498
##     gamma = 0.259593
##     phi   = 0.938777
##
##   Initial states:
##         l        b        s1        s2        s3       s4       s5        s6
##  65.37631 0.260539 0.9915212 0.9146325 0.9317244 1.850616 1.084848 0.9468333
##         s7        s8        s9       s10       s11       s12
##  0.8698946 0.8728443 0.8494383 0.8509356 0.9226715 0.9140402
##
```

```
##   sigma^2:  0.004
##
##      AIC     AICc      BIC
## 4789.884 4791.505 4863.487
```

```
fit %>%
  components(fit)%>%
  autoplot()
```

```
## Warning: Removed 12 row(s) containing missing values (geom_path).
```



Question 3

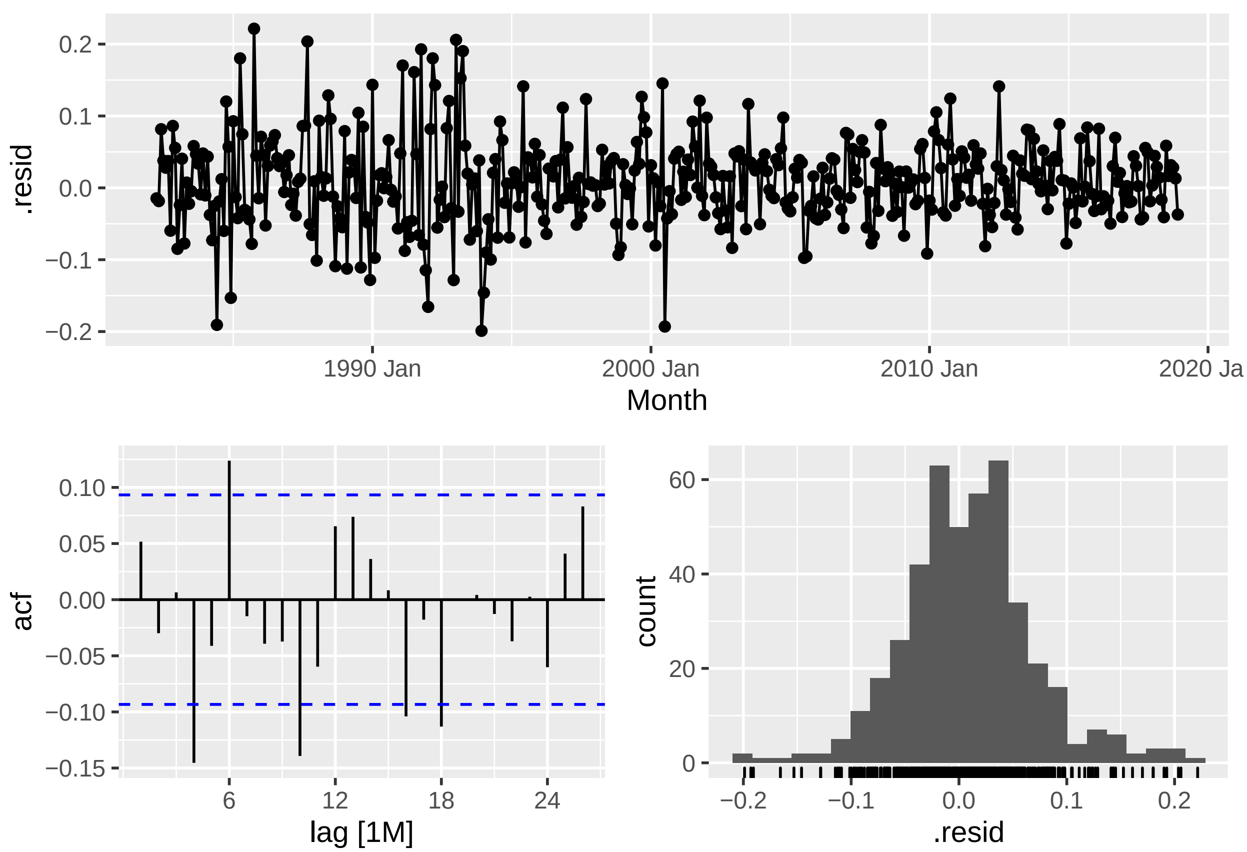
Plot the residuals from the model and comment on these (no more than 50 words). Perform some diagnostic checks and comment on whether you are satisfied with the fit of the model. (Make sure you state all relevant information for any hypothesis test you perform such as the null hypothesis, the degrees of freedom, the decision, etc.). (10 marks)

Significance at lag 4, 6, 10, 16, 18. Shows that it is not a part of a white noise series. There is some correlation present in the residuals. However, it does not fully capture the dynamic in the time series. The error is not homoscedastic, appears to be heteroskedasticity in the first half of the series. Histogram is not normally distributed.

We perform a Ljung-Box test below: H0: That our model does not show lack of fit H1: That our model does show a lack of fit

Degree of freedom = 17 P-value = 2.244262e-08 $<$ 0.05. Therefore, we reject the null hypothesis H0 for all p-values and conclude that the residuals do not have white noise patterns. Additionally, the individual ACF plot for the method is not white noise.

```r
fit %>%
  select(hwdamped) %>%
  gg_tsresiduals()
```




```r
fit %>% tidy()
```

```
## # A tibble: 17 x 3
```

```
##     .model   term   estimate
##     <chr>    <chr>    <dbl>
##   1 hwdamped alpha  0.500
##   2 hwdamped beta   0.00280
##   3 hwdamped gamma  0.260
##   4 hwdamped phi    0.939
##   5 hwdamped l     65.4
##   6 hwdamped b      0.261
##   7 hwdamped s0     0.992
##   8 hwdamped s1     0.915
##   9 hwdamped s2     0.932
##  10 hwdamped s3     1.85
##  11 hwdamped s4     1.08
##  12 hwdamped s5     0.947
##  13 hwdamped s6     0.870
##  14 hwdamped s7     0.873
##  15 hwdamped s8     0.849
##  16 hwdamped s9     0.851
##  17 hwdamped s10    0.923
```

```r
fit %>%
  augment() %>%
  filter(.model=="hwdamped") %>%
  features(.innov, ljung_box, dof=17, lag=24)
```

```
## # A tibble: 1 x 3
##   .model   lb_stat    lb_pvalue
##   <chr>      <dbl>        <dbl>
## 1 hwdamped    49.0 0.0000000224
```

Question 4

Let R select an ETS model. What model has been chosen and how has this model been chosen? (No more than 100 words). (6 marks)

Using the report function, R has selected the [M Ad M] model which is the Holt-Winters damped method. Where the error is multiplicative, the seasonal component is additive damped and the seasonal component is multiplicative.

R has selected this model using the report function as it estimate all the smoothing parameters by maximizing the likelihood and returns information about the fit of the model. For models with multiplicative errors, not equivalent to minimizing SSE, the Nelder - Mead method is used. R selects the model with the lowest (minimizing) Akaike's information criterion corrected because smaller the criterion, better the fit. It also chooses a model with a smaller Bayesian information criterion as likelihood increases.

```r
fit <- my_series %>% model(ETS(y))
report(fit)
```

```
## Series: y
## Model: ETS(M,Ad,M)
##   Smoothing parameters:
##     alpha = 0.4996205
##     beta  = 0.002797498
##     gamma = 0.259593
##     phi   = 0.938777
##
##   Initial states:
##         l        b        s1        s2        s3       s4       s5        s6
##  65.37631 0.260539 0.9915212 0.9146325 0.9317244 1.850616 1.084848 0.9468333
##         s7        s8        s9       s10       s11       s12
##  0.8698946 0.8728443 0.8494383 0.8509356 0.9226715 0.9140402
##
##   sigma^2:  0.004
##
##      AIC     AICc      BIC
## 4789.884 4791.505 4863.487
```

Question 5

Comment on how the model chosen by R is different to the model you have specified (no more than 50 words). Which of the two models would you choose and why? (no more than 50 words). (Hint: think about model selection here but also check your residuals).

If the models are identical specify a plausible alternative. Give a brief justification for your choice (no more than 100 words). (Hint: also check the residuals from this model). (12 marks)

Models are identical, plausible alternative is [M A M], closed AICC & BIC values compared to R recommendation.

The selected model [Ad M] & the chosen model by R [M Ad M] are identical (Holt-winters damped method). A plausible alternative model would be the [M A M] model (Holt-Winters multiplicative method.) This alternative model was chosen because the values for the Akaike's information criterion corrected and Bayesian information criterion are the closest to that of the damped model. This alternative was also chosen because both errors and seasonality are multiplicative but the trend seems linear which is why the none damped model is chosen for comparison.

```
fit1<- my_series %>%
  model(

    hw=ETS(y~error("M")+trend("A")+season("M")),
  )
fit1 %>% glance()
```

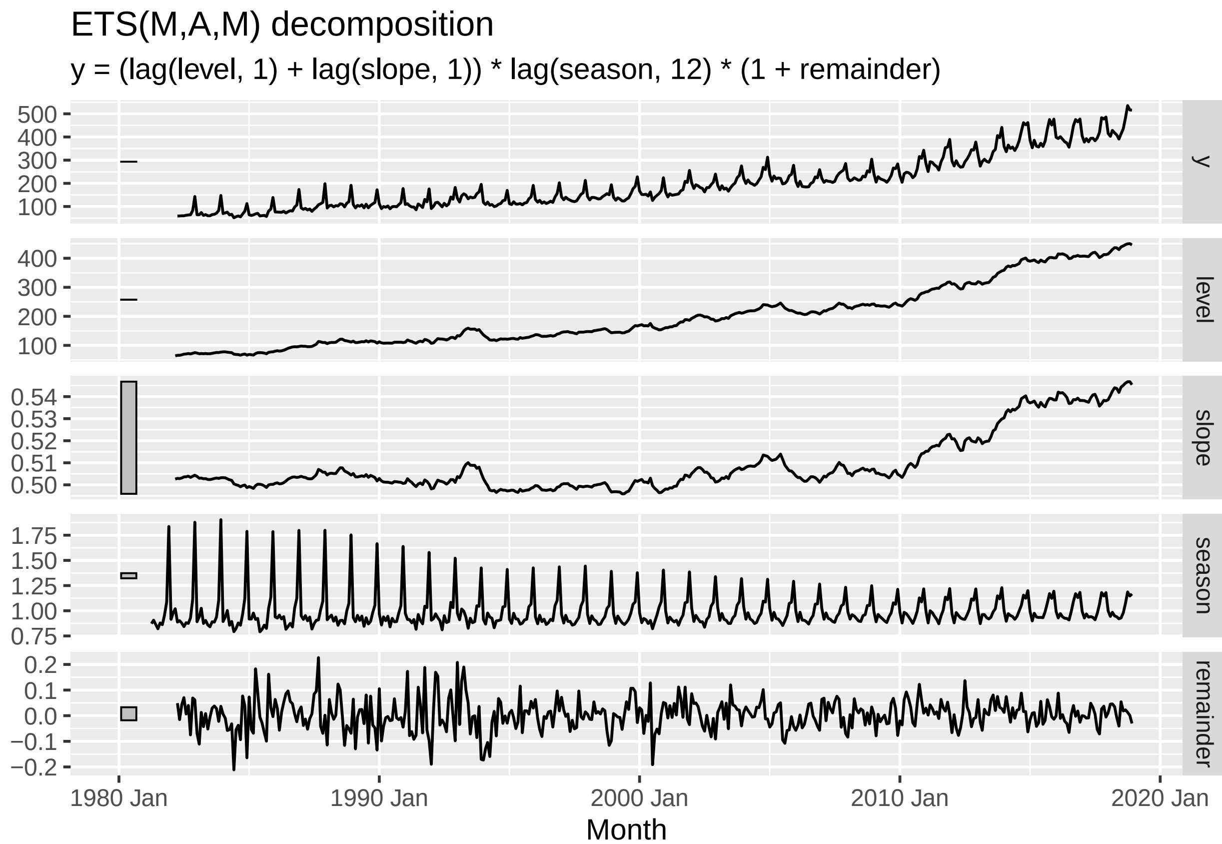
```
## # A tibble: 1 x 9
##   .model  sigma2 log_lik   AIC  AICc   BIC   MSE  AMSE    MAE
##   <chr>    <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  <dbl>
## 1 hw     0.00397  -2380. 4795. 4796. 4864.  118.  154. 0.0462
```

```
fit1 %>%
  report(fit1)
```

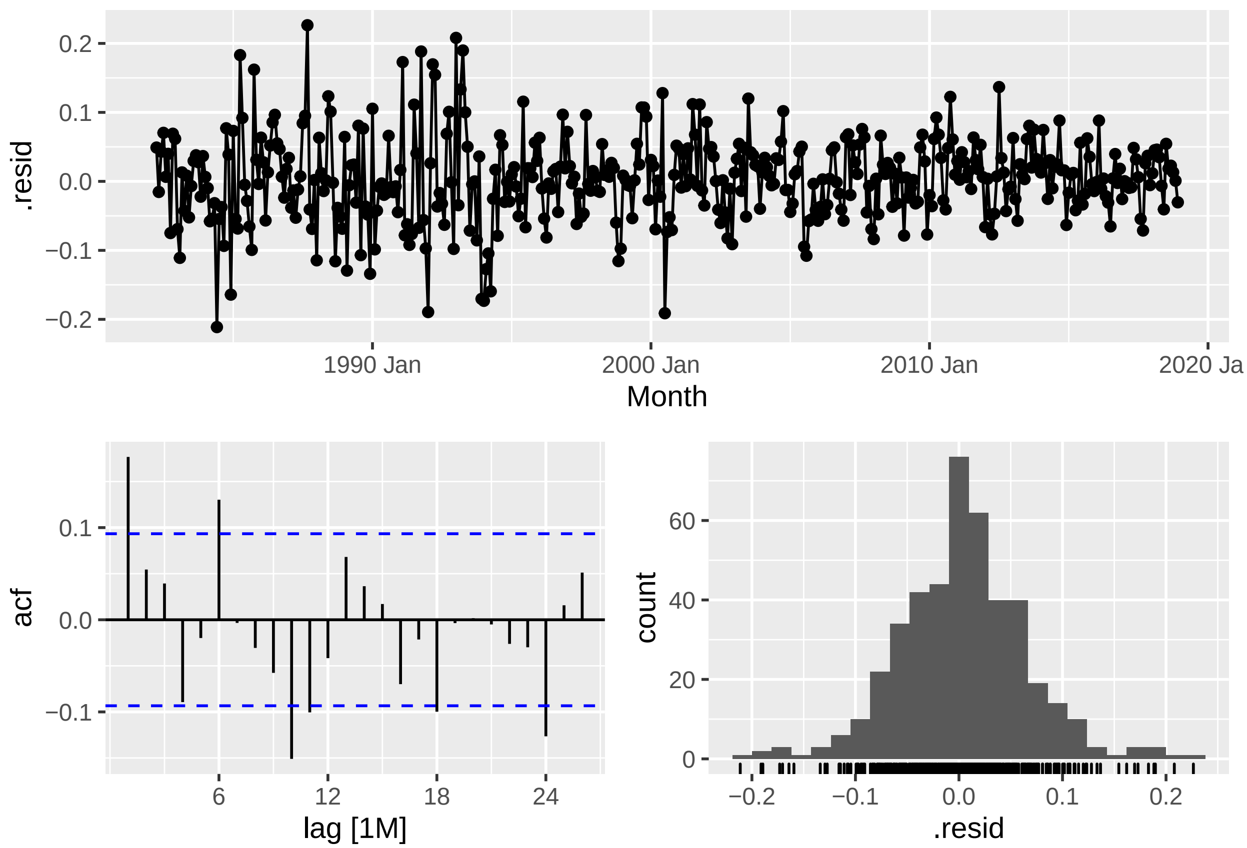
```
## Series: y
## Model: ETS(M,A,M)
##   Smoothing parameters:
##     alpha = 0.3732104
##     beta  = 0.0001016158
##     gamma = 0.3709703
##
##   Initial states:
##         l         b        s1        s2       s3       s4       s5        s6
##  64.10606 0.5026183 1.020175 0.9809043 0.914748 1.836815 1.086988 0.9555013
##         s7      s8        s9       s10       s11       s12
##  0.8619233 0.87922 0.8217221 0.8631075 0.9069135 0.8719828
##
##   sigma^2:  0.004
##
##      AIC     AICc      BIC
## 4794.890 4796.337 4864.404
```

```r
fit1 %>%
  components(fit)%>%
  autoplot()
```

```
## Warning: Removed 12 row(s) containing missing values (geom_path).
```




```r
fit1 %>%
  select(hw) %>%
  gg_tsresiduals()
```



0.2
0.1
0.0
−0.1
−0.2
.resid
1990 Jan
2000 Jan
2010 Jan
2020 Ja
Month
0.1
0.0
−0.1
acf
6
12
18
24
lag [1M]
60
40
20
0
count
−0.2
−0.1
0.0
0.1
0.2
.resid

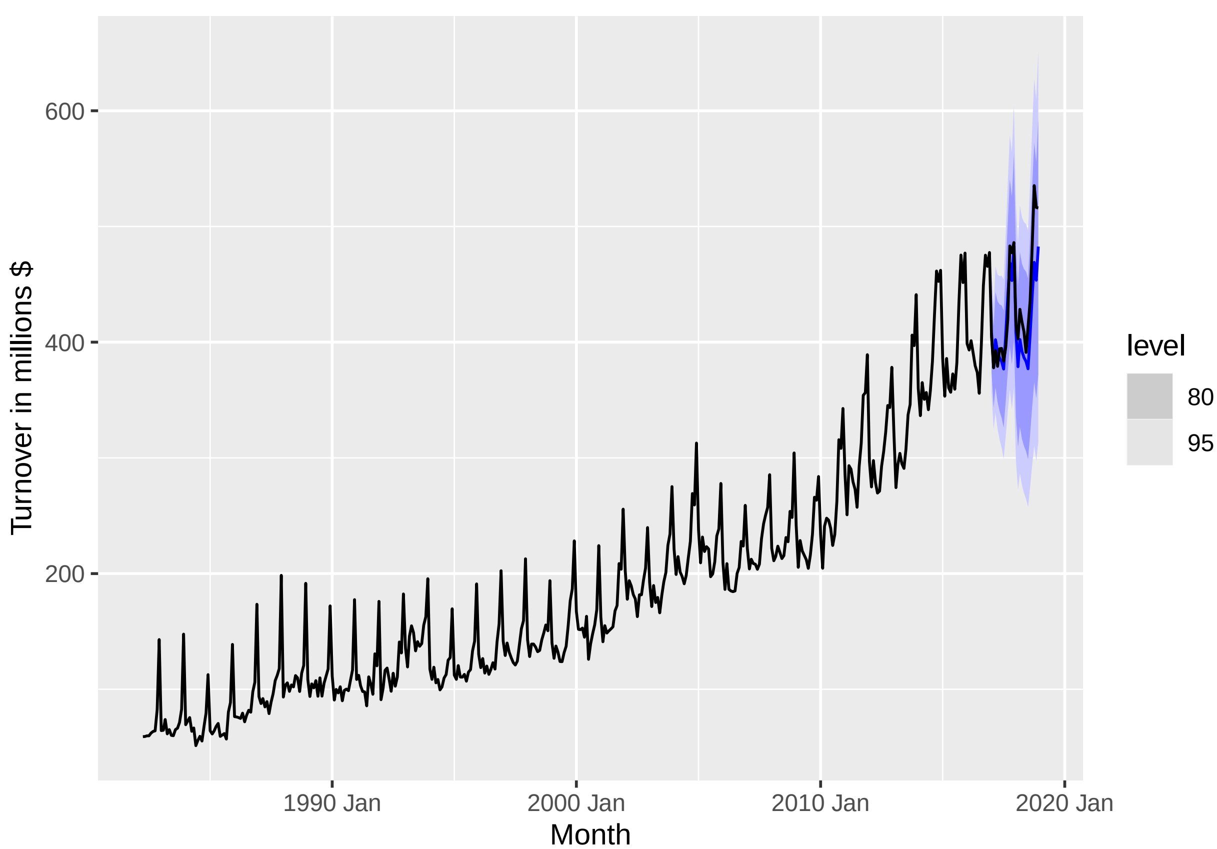
Question 6

Generate forecasts for the last two years of your sample using both alternative ETS models (you will need to re-estimate both models over the appropriate training sample). Plot the forecasts and forecast intervals. Briefly comment on these. Which model does best? (No more than 100 words). (8 marks)

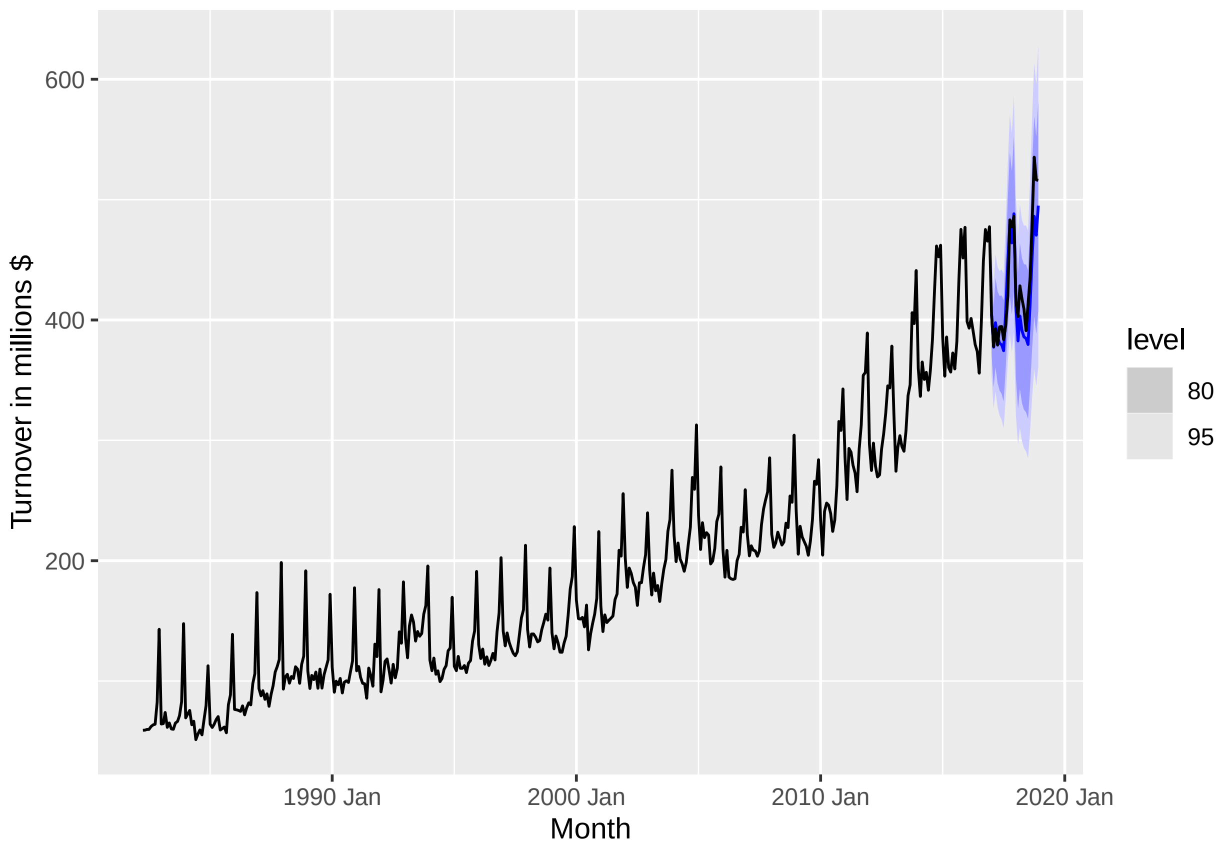
The M Ar M model which R recommends is the best. More accurate point forecast with larger forecast intervals.

The forecast for the last two years using the Holt-Winters damped method shows a slightly larger forecast interval. However, the forecast from Holt-winters damped method is identical to those from the Holt-Winters multiplicative model. Thus, the point forecasts obtained from the method and from the two models that underlie the method are identical since their smoothing parameters are very similar. For multiplicative models such as these, the point forecasts do not have a mean equal to the forecast distribution. Since both models are multiplicative, the prediction intervals will be similar.

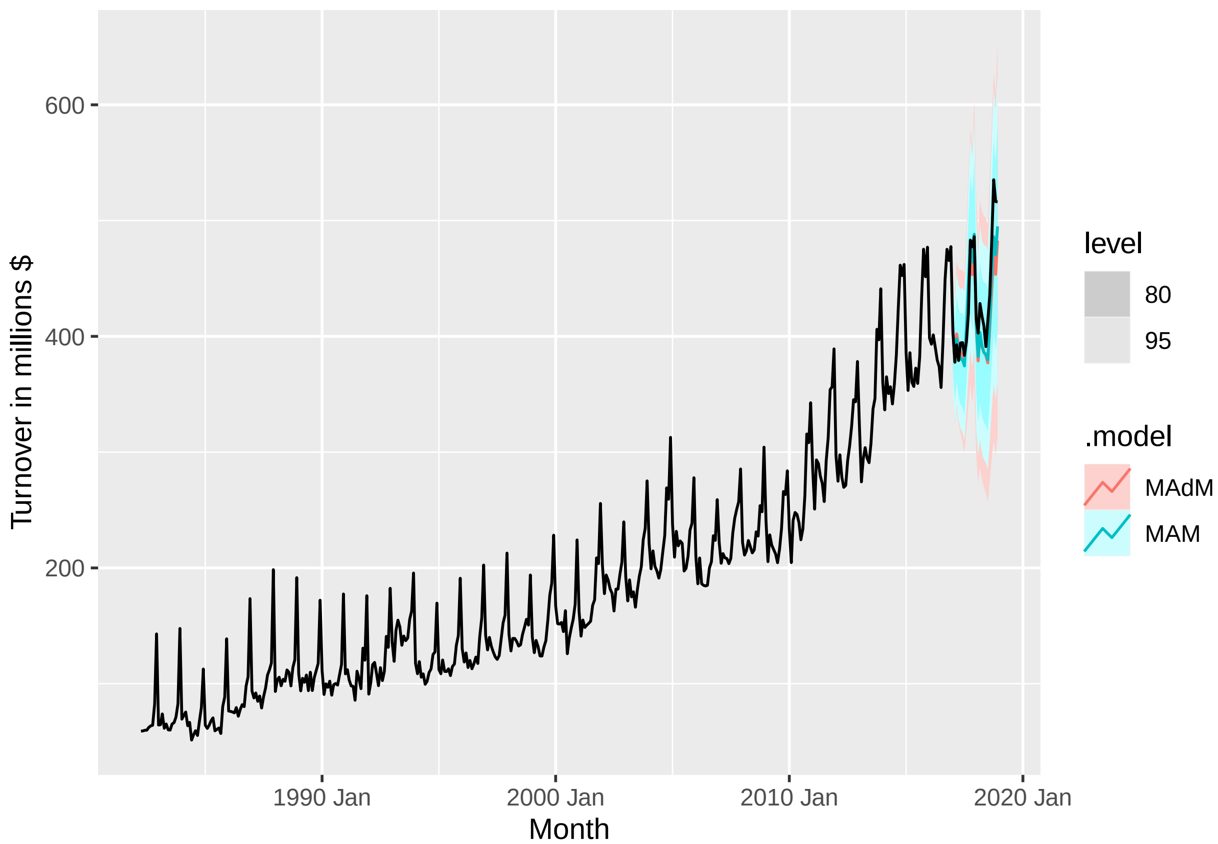
```
train <- my_series %>%
  slice(1:(n() - 24))


test <- my_series %>%
  slice((n() - 23):n())

train %>%
  model(hwdamped= ETS(y~error("M")+trend("Ad")+season("M")),)%>%
  forecast (h="2 years") %>%
  autoplot() + autolayer(my_series, y)+ labs(y="Turnover in millions $")
```



```
train %>%
  model(hw= ETS(y~error("M")+trend("A")+season("M")),)%>%
  forecast (h="2 years") %>%
  autoplot() + autolayer(my_series, y)+ labs(y="Turnover in millions $")
```




```
train %>%
  model(MAdM= ETS(y~error("M")+trend("Ad")+season("M")),MAM= ETS(y~error("M")+trend("A")+season("M")))%
  forecast (h="2 years") %>%
  autoplot() + autolayer(my_series, y)+ labs(y="Turnover in millions $")
```


Turnover in millions $
600
400
200
1990 Jan
2000 Jan
2010 Jan
2020 Jan
Month
level
80
95
.model
MAdM
MAM

Question 7

Generate forecasts for the two years following the end of your sample using your chosen model. Plot them and briefly comment on these. (No more than 100 words). (4 marks)

The forecasts for the two years using the Holt-Winters damped method is shown. What is interesting is that the 80% level shown is also represented by the Hold-Winters multiplicative model. This tells us that The Holt-Winters damped method is much more accurate in forecasting but has a larger forecast interval.

```
train %>%
  model(hwdamped= ETS(y~error("M")+trend("Ad")+season("M")),)%>%
  forecast (h="2 years") %>%
  autoplot() + autolayer(my_series, y)+ labs(y="Turnover in millions $")
```