Code ▾

Assignment 3: Suggested Solutions & Feedback

ETF3231/5231

Read and tidy up the data using previous code.

> For the tasks that follow you will be modelling the original data
> without any transformation 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)

Hide

```
myts %>% autoplot(y) +
  labs(title ="Add you title here",
       lab="add y label")
```
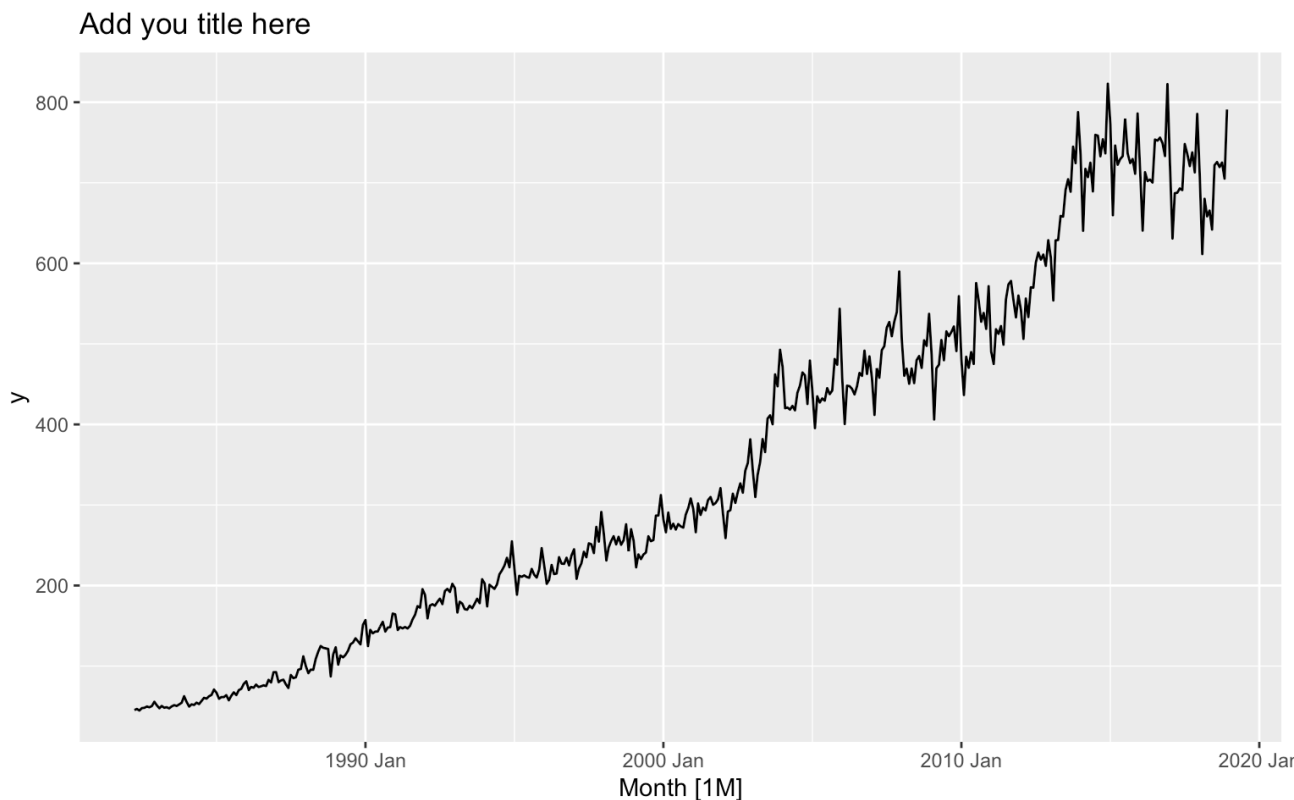
Marking guide:

- Plot: labels, units and heading. (1m)
- Choose an ETS model (1m)
- Justify the choice for trend (damped or not). (2m)
- Justify the choice for seasonal. (2m)
- Justify the choice for error. (2m)
- -2: Exceeding word limit.

Expectation:

- Plot the time series (with labels, units and heading).
- Select an ETS model.
- Justify your choice for trend (damped or not), seasonal and error components.

Sample answer:

If your model was ETS(M,Ad,M) then you would justify your choices as follows:

- Data seems to be trending, so selecting a trend seems to be appropriate.
- The trend seems fairly flat towards the end, so a damped trend is selected.
- Multiplicative seasonality is used because the seasonal variation increases in size as the level of the series increases.
- Multiplicative errors fit better with multiplicative seasonality - remember the combinations of ETS(A,*,M) are the ones that can potentially cause estimation trouble and were labelled as forbidden/unstable.

- In some assignments particular model components were stated, but not described or, vice versa, described, but not named in terms of the ETS model. Many suggested models were taking additive seasonal component instead of multiplicative one required which was more obvious from the plot.

Common mistakes:

- decompose the series is not needed for the justification of the model
- do not attempt to run the ETS model, use own judgement by looking at the time series itself.

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)

Hide

```
ETS_MAdM <- myts %>% model(ETS(y ~ error("M")+trend("Ad")+season("
M")))

options(digits=4)

ETS_MAdM %>% report()
```

```
## Series: y
## Model: ETS(M,Ad,M)
##    Smoothing parameters:
##      alpha = 0.6655
##      beta  = 0.02106
##      gamma = 0.000107
##      phi   = 0.98
##
##    Initial states:
##    l[0]   b[0]    s[0]  s[-1] s[-2] s[-3] s[-4] s[-5] s[-6] s[-7] s
[-8]   s[-9]
##   46.54 0.1436 0.9789 0.9039 1.032  1.11 1.005 1.038 1.002 1.014 1.
015 0.9586
##   s[-10] s[-11]
##   0.9805 0.9629
##
##    sigma^2:  0.0017
##
##   AIC AICc  BIC
## 4811 4813 4885
```

Hide

```
#or
ETS_MAdM %>% tidy() %>% select(-.model)
```

```
## # A tibble: 18 × 2
##    term     estimate
##    <chr>       <dbl>
##  1 alpha     0.666
##  2 beta      0.0211
##  3 gamma     0.000107
##  4 phi       0.980
##  5 l[0]     46.5
##  6 b[0]      0.144
##  7 s[0]      0.979
##  8 s[-1]     0.904
##  9 s[-2]     1.03
## 10 s[-3]     1.11
## 11 s[-4]     1.00
## 12 s[-5]     1.04
## 13 s[-6]     1.00
## 14 s[-7]     1.01
## 15 s[-8]     1.02
## 16 s[-9]     0.959
## 17 s[-10]    0.980
## 18 s[-11]    0.963
```
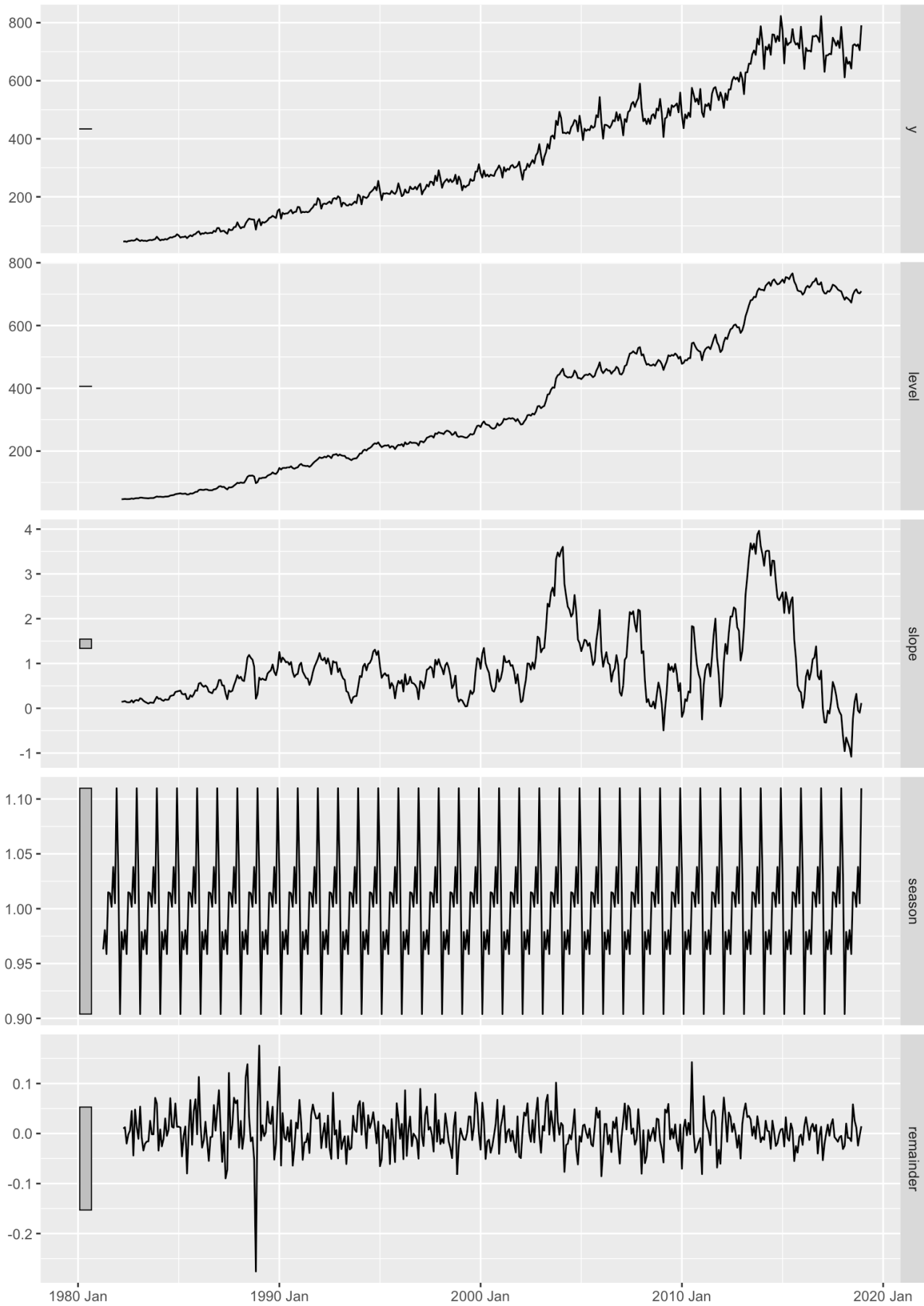
Hide

```
ETS_MAdM %>% components() %>%  autoplot()
```

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

ETS(M,Ad,M) decomposition

y = (lag(level, 1) + 0.979998847090223 * lag(slope, 1)) * lag(season, 12) * (1 + remainder)

Marking guide:

- Estimated model output (3m)
- Plot the estimated states (2m)
- Recognise that initial states are also estimated. (1m)
- Comment on the smoothing parameters and relate the estimated smoothing coefficients with the plot (8m = 2 + 3 + 3 for each level, trend, seas)
- -1: Exceeding word limit.

Expectation:

- Show the estimated model output.
- Comment on the estimated smoothing parameters.
- Recognise that initial states are also estimated.
- Plot the estimated states.
- Comment on how the estimated smoothing coefficients relate to plots these.

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)

Hide

```
ETS_MAdM %>% gg_tsresiduals()
```
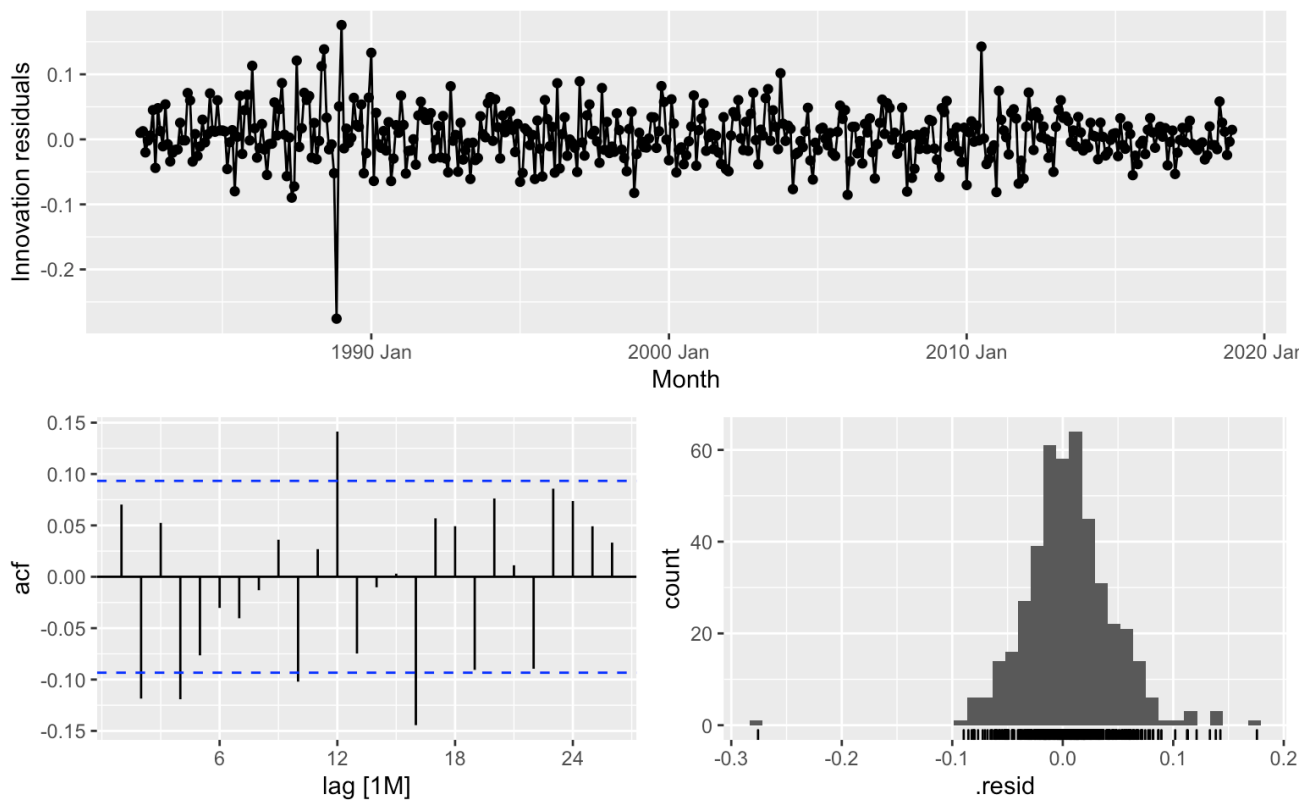
Hide

```
augment(ETS_MAdM) %>% features(.innov, ljung_box, lag=24, dof=17)
```

```
## # A tibble: 1 × 3
##    .model                                                    lb_stat
lb_pvalue
##    <chr>                                                       <dbl>
<dbl>
## 1 "ETS(y ~ error(\"M\") + trend(\"Ad\") + season(\"M\"))"      65.7
1.09e-11
```

Marking guide:

- Plot/Comment on the residuals (2m)
- Plot/Comment on the histogram of the residuals (2m)
- Plot/Comment on the ACF plot (2m)
- Perform LB test (1m), with the hypothesis (1m) and correct df (1m) and comment on the result (1m)
- -1: Exceeding word limit.

Expectation:

- Plot the residuals and comment.
- Plot the histogram and comment.

- Plot the ACF and comment.
- State the null hypothesis and perform a Ljung-Box test and comment.
- Comment on whether you are satisfied with the fit of the model.

Common errors:

- Ho: The residuals are white noise or more formally $\rho_1 = \cdots = \rho_k = 0$.

- Often the decision rule was not stated clearly (without a comparison of p-value with the level of significance). Null hypothesis statements were missing or there were mistakes in the conclusion after the hypothesis testing.

- Many students did not set the correct values for lag and dof arguments in the Ljung-Box test.

    - For seasonal data, set lag = $2 \times m = 2 \times 12 = 24$, where m = seasonal period.
    - For non-seasonal data, set lag $= 10$. These are of course a subjective choice and can be different.
- `dof` is lag chosen minus the number of "free" estimated parameters. (Remember for seasonal initial values one parameter is not free). Note that dof=17 for the model selected above. You must set the correct dof for the ETS model you selected in question 1.

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)

Hide

```
ETS_auto <- myts %>%  model(ETS(y))

ETS_auto %>% report()
```

```
## Series: y
## Model: ETS(M,Ad,M)
##   Smoothing parameters:
##     alpha = 0.6655
##     beta  = 0.02106
##     gamma = 0.000107
##     phi   = 0.98
##
##   Initial states:
##   l[0]   b[0]   s[0]  s[-1] s[-2] s[-3] s[-4] s[-5] s[-6] s[-7] s
[-8]   s[-9]
##  46.54 0.1436 0.9789 0.9039 1.032  1.11 1.005 1.038 1.002 1.014 1.
015 0.9586
##  s[-10] s[-11]
##  0.9805 0.9629
##
##   sigma^2:  0.0017
##
##  AIC AICc  BIC
## 4811 4813 4885
```

Marking guide:

- State the chosen model (2m)
- Explain how the model is obtained (1m)
- The model is selected based on smallest AICc (2m) and excludes unstable models (1m)

Expectation:

- State the model.

- How has this model been chosen? The ets function returns the model with the smallest AICc.

- Without restrictions, the ETS() function fits all possible combinations and returns the model with the smallest AICc. However, the ETS function is restricted by default and excludes unstable models.

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)

Marking guide:

- Comment on how two models are different (different alpha, gamma or other characteristics. (6m)
- State the selected model and why (min AICc). (1m)
- Discuss the chosen model (5m)
  - visualise residual plot 1
  - comment the residual plot 1
  - ACF 1
  - histogram 1
  - LB test 1
- state the predictive accuracy (1m)
- -1: Exceeding word limit
- -0.5: Incorrect interpretation

Detail breakdown:
0-1.5 -> no explanantion
2 -> Unaccecptable explanation
3 -> Inadequate explanation
4 -> Interesting explanation
+2 -> with the comment upon coefficients

Expectation if the models are different:

- Comment on how two models are different.
- The chosen model.
- The reason for choosing the best model.
- Check the residuals (a brief discussion on the residual plot and ACF – histogram and Ljung-Box).

Expectation if the models are identical:

- Specify a plausible alternative.
- The chosen model.
- The reason for choosing the best model.
- Check the residuals (a brief discussion on the residual plot and ACF – histogram and Ljung-Box test).

Given the severe flattening of the trend in the last few years a plausible alternative is:

Hide

```
ETS_alt <- myts %>%  model(ETS(y ~ error("M")+trend("N")+season("
M")))
ETS_alt %>% report()
```

```
## Series: y
## Model: ETS(M,N,M)
##   Smoothing parameters:
##     alpha = 0.5946
##     gamma = 0.253
##
##   Initial states:
##   l[0]   s[0]  s[-1] s[-2] s[-3]  s[-4] s[-5] s[-6]  s[-7]  s[-8]
s[-9] s[-10]
##  54.15 0.9731 0.9699 1.106  1.16 0.9737 1.019 1.003 0.9853 0.9554
0.8901 0.9927
##  s[-11]
##  0.9714
##
##   sigma^2:  0.0021
##
##  AIC AICc  BIC
## 4884 4885 4945
```

Which of the two models would you choose?

- This is not a subjective choice. Select the model with the minimum AICc rather than the one with the minimum forecast accuracy measures.
- Since both models are in the same ETS family, AICc can be used to compare the models. Remember the advantage of using an information criterion such as the AICc is that it uses all the data and not just a small test sample that the forecast accuracy measures such as RMSE (unless something has gone obviously wrong which is rarely the case).

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)
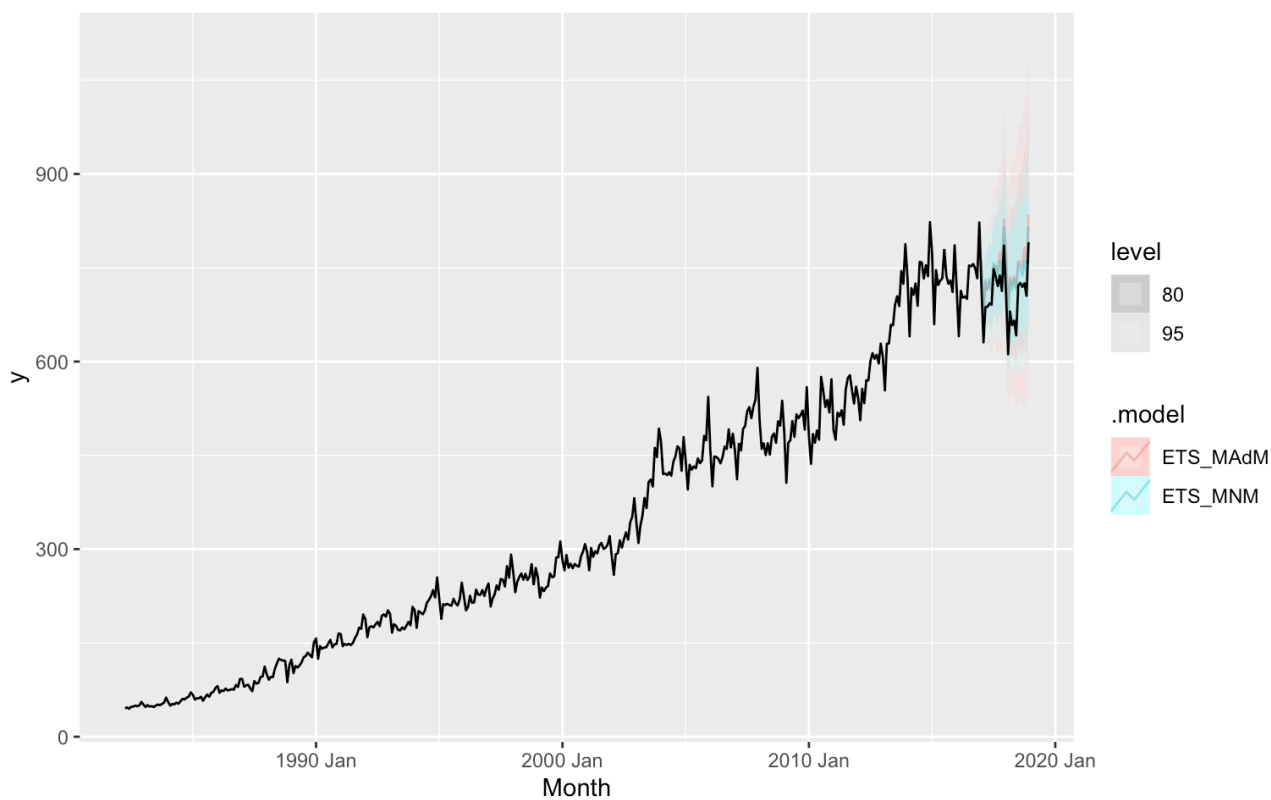
Hide

```
myts_train <- myts %>%
  slice(1:(n() - 24))

ETS_train_fit <- myts_train %>%
  model(ETS_MAdM=ETS(y~error("M")+trend("Ad")+season("M")),
        ETS_MNM=ETS(y~error("M")+trend("N")+season("M"))
)
ETS_train_fit %>% tidy() %>%
  pivot_wider(names_from=.model,values_from=estimate)
```

```
## # A tibble: 18 × 3
##    term     ETS_MAdM ETS_MNM
##    <chr>       <dbl>   <dbl>
##  1 alpha     0.607    0.481
##  2 beta      0.0166   NA
##  3 gamma     0.000100 0.0312
##  4 phi       0.979    NA
##  5 l[0]     46.5     53.6
##  6 b[0]      0.402    NA
##  7 s[0]      0.981    0.987
##  8 s[-1]     0.909    0.905
##  9 s[-2]     1.03     1.03
## 10 s[-3]     1.11     1.10
## 11 s[-4]     1.01     1.02
## 12 s[-5]     1.04     1.02
## 13 s[-6]     0.999    0.985
## 14 s[-7]     1.01     1.01
## 15 s[-8]     1.01     1.02
## 16 s[-9]     0.958    0.963
## 17 s[-10]    0.981    1.00
## 18 s[-11]    0.964    0.963
```

Hide

```
fc <- ETS_train_fit %>% forecast(h=24)
fc %>% autoplot(myts, alpha=0.4)
```

Hide

```
fc %>% autoplot(filter(myts, year(Month)>=2015), alpha=0.6)
```
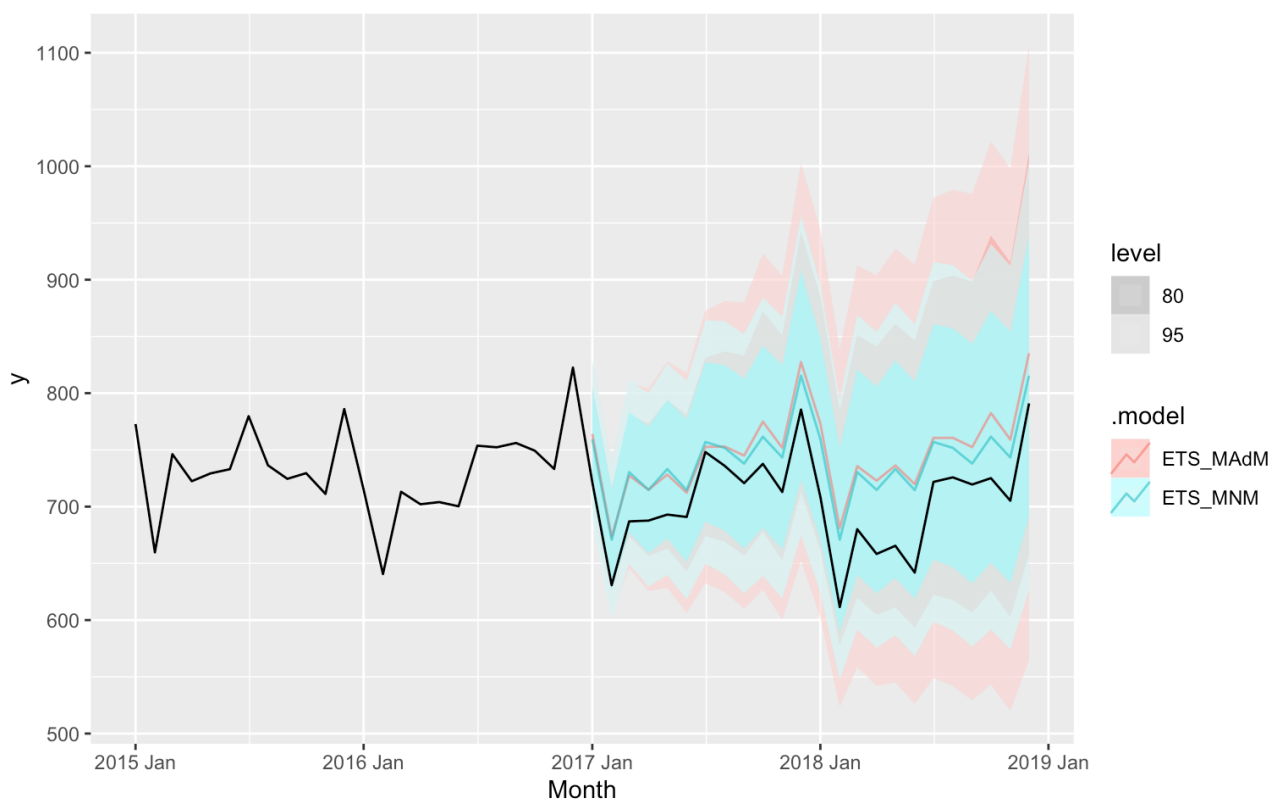

Hide

```
fc %>% accuracy(myts) %>% select(.type,ME, RMSE, MAPE, MASE)
```

```
## # A tibble: 2 × 5
##   .type    ME  RMSE  MAPE  MASE
##   <chr> <dbl> <dbl> <dbl> <dbl>
## 1 Test  -43.2  46.8  6.25  1.64
## 2 Test  -36.4  39.9  5.29  1.38
```

Comment on the results.

Marking guide:

- Plot the forecasts from both models over the data (with forecast intervals, for h=24). Zoom in if required or make sure visualisation is good enough (4m)
  - Slice the data 1m (if not 0)
  - Plot the forecasts from both models over the data 2m (if not 0, 1 if seperate plot)
  - Zoom in 1m (if not 0)
- Comment on the forecasts (2m)
- Comment on the forecast interval (2m)
- Predictive accuracy (printout +1m) + (interpret +1m) NO RMSE -1
- Bonus 2 marks if test sample comparison is performed and commented upon (-1 if MASE is not showing).

Expectation:

- Plot the forecasts as a continuation of the data.
- Comment on the forecasts. Are the forecasts plausible?
- Comment on the forecast intervals. How wide are the intervals?

Some students did not plot the forecast intervals and did not comment on them. It was clear in the question that you were asked to produce a graph with two sets of forecasts and forecast intervals, and comment on these.

Common errors:

- The explanation on prediction interval are wider for year 2018 compared to year 2017 will not be accepted as the comment on the forecast interval. As the time horizon gets larger, the forecast interval will be wider as well with some exceptional cases.
- The explanation on 95% prediction interval is wider than 80% is not accepted as a comment as well as this is a fact.
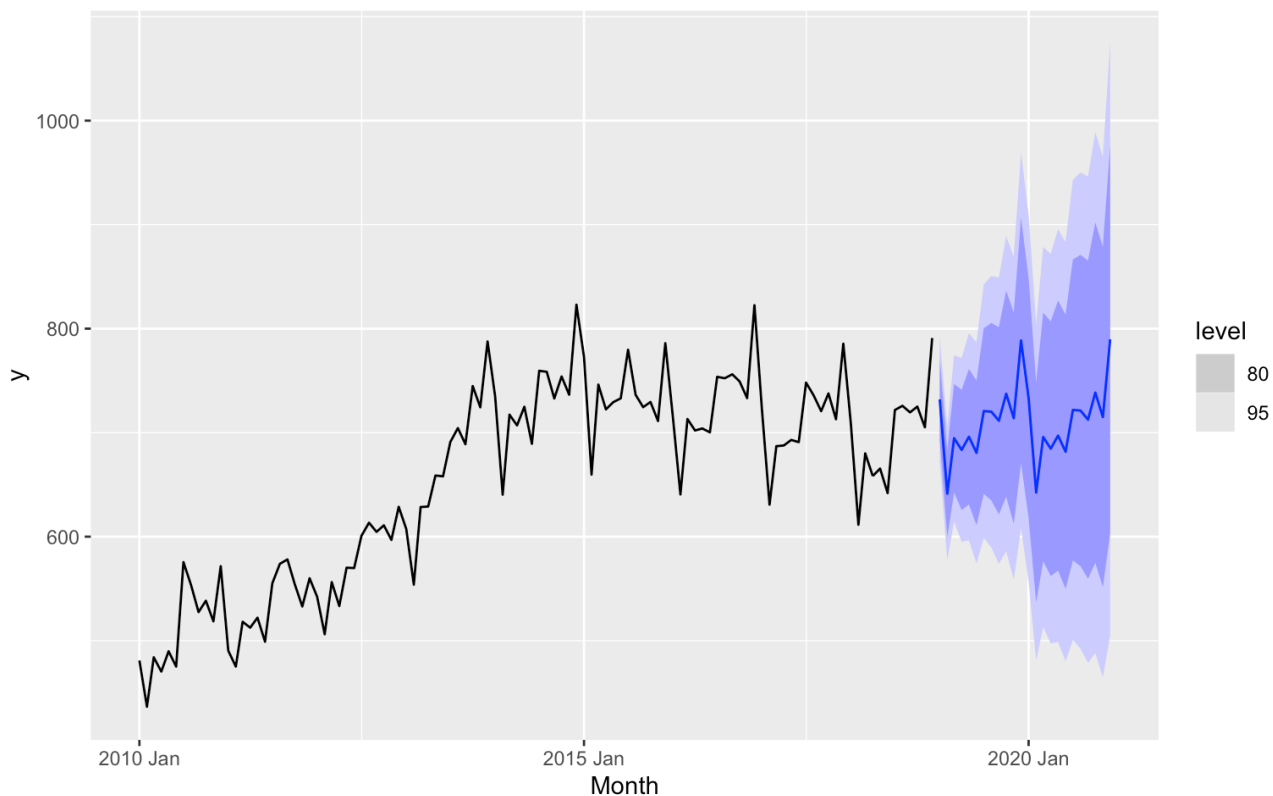
Question 7

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)

Hide

```
ETS_MAdM %>% forecast() %>% autoplot(myts)
```
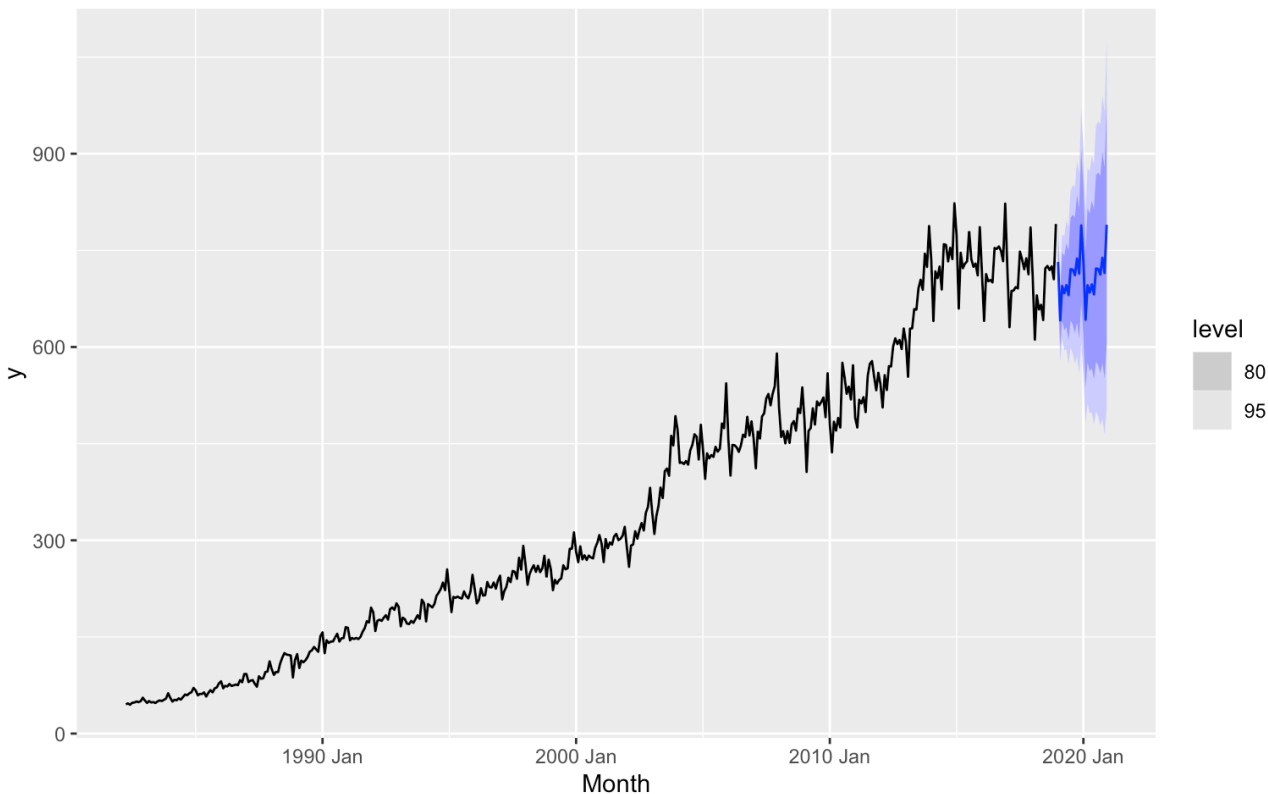

Hide

```
ETS_MAdM %>% forecast() %>% autoplot(filter(myts, year(Month)>=2010))
```

Marks Breakdown:

- Plot of point forecast for h=24 in original unit. (2m)
- Comment on the forecasts and intervals (the width for my case here is important given the uncertainty). (2m)
- REMEMBER: Minus 1 overall if exceeding word limits.

Expectation:

- Plot the forecasts as a continuation of the data.
- Plot the forecasts on the original scale.
- Comment on the forecasts. Do they look plausible?

Common mistake:
* Most of the explanation on prediction interval are due to the increase in time horizon.
* Should relate the prediction interval with the potential left over pattern in the residual analysis.