Predictive Analytics - Exercise Sheet 2

(graded)

Prof. Dr. Benjamin Buchwitz

2024

Task 1

The idea of exponential smoothing is illustrated by equation 1 which has been extracted from Hyndman and Athanasopoulos (2021) (subsection 8.1).

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + ... \tag{1}$$

Based on equation 1 answer the following questions and do the following:

a) Which factors embody the weights of exponential smoothing?

b) What is the meaning of α?

c) Write a function that computes the weights (see task 1.a)) for the last T observations dependent on α. Choose $T = 20$ and different values for α: $\alpha_1 = 0.2$, $\alpha_2 = 0.4$, $\alpha_3 = 0.6$ and $\alpha_4 = 0.8$ and plot the resulting values (curves) in one plot. Add a legend.

d) What can you see in your created plot from task 1.c)?

Task 2

a) Write down the forecast equation $\hat{y}_{T+h|T}$ of

   a.1) Holt's linear trend method and

   a.2) Holt's linear trend method with a damping parameter $0 < \phi < 1$.

b) Based on the two forecast equations from task 2.a) produce forecasts for the next $h = 50$ time points: Assume that $l_T = 25$, $b_T = 0.33$ and $\phi = 0.88$. Implement a smart and short solution since the way of your implementation will also be graded!
Plot your forecasts based on the two methods in one plot.

   b.1) Describe and compare the produced forecasts based on the two forecast equations.

   b.2) What happens when you increase the value of ϕ? Illustrate it by choosing at least three different values for ϕ.

Task 3

a) Write down the equations of the Holt-Winters' additive method (including the forecast equation).

b) What does the index of the seasonal component $t + h - m(k + 1)$ in the context of computing the

   b.1) fitted values $\hat{y}_{t+h|t}$ and b.2) forecasts $\hat{y}_{T+h|T}$

indicate?

c) After applying the Holt-Winters' additive method to the Australian domestic tourism data, the following values are obtained based on $\alpha = 0.2620$, $\beta^* = 0.1646$ and $\gamma = 0.0001$ (see subsection 8.3 in Hyndman and Athanasopoulos (2021) for complete information):

Table 1: Resulting values after applying the Holt-Winters' additive method to the Australian domestic tourism data from Hyndman and Athanasopoulos (2021).

Quarter	**Time**	**Observation**	**Level**	**Slope**	**Season**	**Forecast**
	t	y_t	l_t	b_t	s_t	$\hat{y}_{t+1\|t}$
1997 Q1	0				1.5	
1997 Q2	1				-0.3	
1997 Q3	2				-0.7	
1997 Q4	3		9.8	0.0	-0.5	
1998 Q1	4	11.8	9.9	0.0	1.5	11.3
1998 Q2	5	9.3	9.9	0.0	-0.3	9.7
1998 Q3	6	8.6	9.7	-0.0	-0.7	9.2
1998 Q4	7	9.3	9.8	0.0	-0.5	9.2
	...	...	...	...	...	...
2017 Q1	80	12.4	10.9	0.1	1.5	12.3
2017 Q2	81	10.5	10.9	0.1	-0.3	10.7
2017 Q3	82	10.5	11.0	0.1	-0.7	10.3
2017 Q4	83	11.2	11.3	0.1	-0.5	10.6
	h					$\hat{y}_{T+h\|T}$
2018 Q1	1					12.9
2018 Q2	2					11.2
2018 Q3	3					11.0
2018 Q4	4					11.2
2019 Q1	5					13.4

Understand how the values in table 1 have been computed. Calculate the first four fitted values by hand and show how the corresponding level, trend and seasonal component have been calculated. How can you reproduce the forecasts for $h = 1, ..., 5$? Calculate them also by hand.

Task 4

Why is it not possible to generate point forecasts **and** prediction intervals based on exponential smoothing methods? Why are statistical models needed? Explain it in your own words!

Task 5

a) How does the Holt's linear trend method look like? Write down its equations (including the forecast equation).

b) Show that the Holt's linear trend method with additive errors results in the following ETS(A,A,N) model:

$y_t = l_{t-1} + b_{t-1} + \epsilon_t$

$l_t = l_{t-1} + b_{t-1} + \alpha\epsilon_t$

$b_t = b_{t-1} + \beta\epsilon_t$

with $\beta = \alpha\beta^*$.
Write down **all** your intermediate steps and name the different steps! *Hint*: Consider how the ETS(A,N,N) model has been constructed based on simple exponential smoothing with additive errors in subsection 8.5 in Hyndman and Athanasopoulos (2021).

Task 6

a) Write your own function to implement Holt's linear trend method (see Task 5.a)) and to make forecasts. The function should take the following arguments:

- `y` (the time series),
- `alpha` (the smoothing parameter α),
- `beta` (the smoothing parameter β^*),
- `level` (the initial level l_0),
- `trend` (the initial trend b_0) and
- `h` (the number of forecasts)

and it should return the following computed values:

- `level_comp` (the level component l_t),
- `trend_comp` (the trend component b_t (slope)),
- `y_fitted` (the fitted values (forecasts) $\hat{y}_{t+1|t}$) and
- `y_h` (the *true* forecasts $\hat{y}_{T+h|T}$).

b) Apply your function from task 6.a) to the following Australian population data (see code chunk below) which is based on the data set `tsibbledata::global_economy` (O'Hara-Wild et al. 2022) and which has been extracted from Hyndman and Athanasopoulos (2021) (subsection 8.2).
Your function should return the forecasts for the **next ten** observations. Does it give the same forecasts as the defined ETS model (based on `fable::ETS()` (O'Hara-Wild, Hyndman, and Wang 2021)) in the code chunk below?
As initial values for α, l_0 and b_0 take the ones which are returned by applying the predefined `ETS`-model (see below) to `aus_economy`. Which value do have to take for β^* and why? Round the initial values to four decimal places.

Visualize the time series `aus_economy`, the return values of your function ($\hat{y}_{t+1|t}$ and $\hat{y}_{T+h|T}$) and the corresponding values which result from applying `ETS()` in **one** plot. Do you expect differences between the return values of your function and the ones resulting from `ETS()`? If yes, which ones?

```
aus_economy <- global_economy %>%
filter(Code == "AUS") %>%
mutate(Pop = Population / 1e6)

fit <- aus_economy %>%
model(
  AAN = ETS(Pop ~ error("A") + trend("A") + season("N"))
)
fc <- fit %>% forecast(h = 10)
```

Task 7

Consider the number of pigs slaughtered in New South Wales named `pigs` from `tsibbledata::aus_livestock` (O'Hara-Wild et al. 2022) (see code chunk below).

The data set `pigs` has been splitted in a training set of 486 observations and a test set of 72 observations (6 years):

```
pigs <- aus_livestock %>% dplyr::filter(State == "New South Wales", Animal == "Pigs")
train_pigs <- pigs[1:486,] # training set of the first 486 observations
test_pigs <- pigs[487:nrow(pigs),] # test set of 72 observations
```

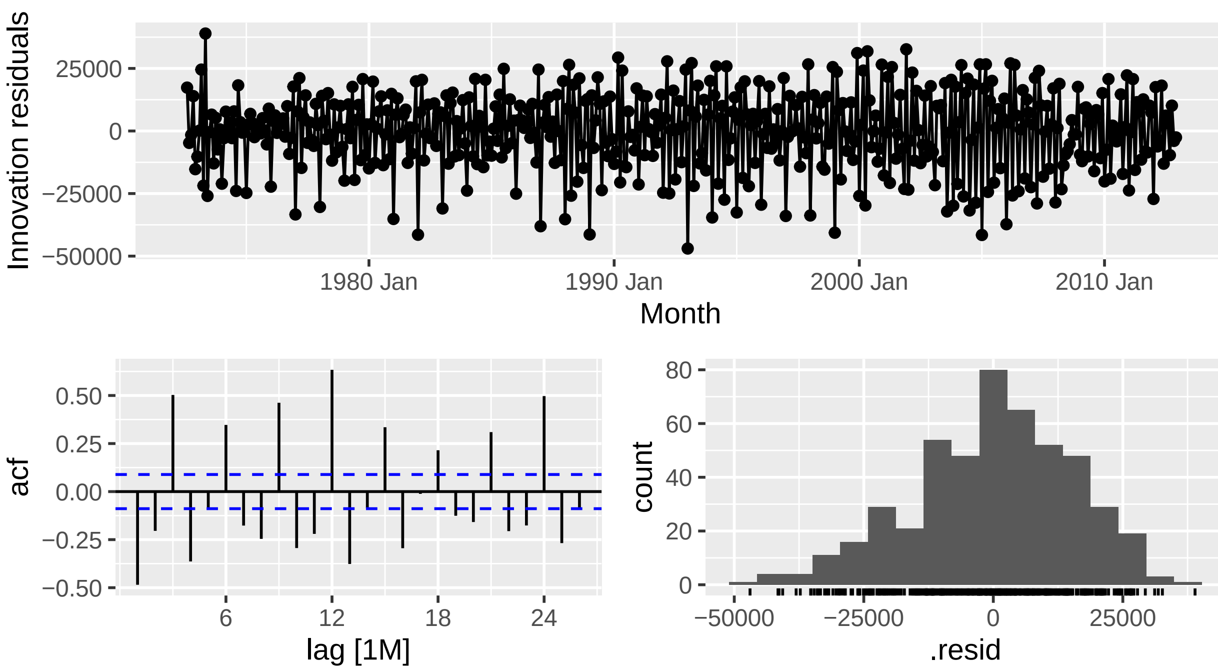
a) Visualize the time series `pigs` and describe it. Do you recognize a trend at the most current observations?

b) Apply an ETS model with additive errors, an additive trend and no seasonality to `pigs`, produce forecasts based on the training set for the test set. Visualize the time series `pigs` and your fitted values and forecasts (with prediction intervals of 80 % and 95 % coverage probability) in **one** plot.

c) Why is the defined ETS model in task 7.b) not the best one?

d) Which ETS model is *best* and why? Give an explanation and plot `pigs` and your fitted values and forecasts based on the chosen ETS model in **one** plot.

e) Check out the residuals of your chosen ETS-model from task 7.d) by plotting a time series, a histogram and an acf plot of the residuals! Describe what you can see in the single plots and interpret them. Then, compare these plots with the following ones that were obtained after applying the drift method to `train_pigs`:



f) Compute the RMSE, MAE, MAPE and MASE after producing forecasts for the test set based on your chosen ETS model from task 7.d). Then, compare these accuracy measures with the following ones that were obtained after applying the drift method to `pigs`. What can you conclude from it?

```
## # A tibble: 1 x 5
##   .model  RMSE   MAE  MAPE  MASE
##   <chr>  <dbl> <dbl> <dbl> <dbl>
## 1 Drift  8091. 6967.  10.1 0.657
```

References

Hyndman, Rob J, and George Athanasopoulos. 2021. *Forecasting: Principles and Practice*. 3rd ed. Springer-Lehrbuch. Melbourne, Australia: OTexts.

O'Hara-Wild, Mitchell, Rob Hyndman, and Earo Wang. 2021. *Fable: Forecasting Models for Tidy Time Series*. https://CRAN.R-project.org/package=fable.

O'Hara-Wild, Mitchell, Rob Hyndman, Earo Wang, and Rakshitha Godahewa. 2022. *Tsibbledata: Diverse Datasets for 'Tsibble'*. https://CRAN.R-project.org/package=tsibbledata.