Predictive Analytics - Exercise Sheet 0

(non-graded)

Prof. Dr. Benjamin Buchwitz

2024

1 Task 1

In the following, the data set `fhswf::whodata` (Buchwitz 2022) with information about the development of different countries will be considered.

a) Install the `fhswf` package:

```
remotes::install_github("bchwtz/fhswf") # 1 P
```

b) Select the variables `country_code`, `year` and `population` from the `whodata` data set, then convert it to a `tsibble` (R package `tsibble` (Wang, Cook, and Hyndman 2020)) with year as the index and country code as the key.

c) What does the key in the created `tsibble` object indicate in general and here in this example?

d) Plot the population by the countries which are indicated by AUS, BRA, DNK, DEU and CZE in the time span from 2010 to 2019.

2 Task 2

a) Filter the time series "Total Private" `Employed` from `fpp3::us_employment` (R. Hyndman 2021) and save it as `us_empl_private`.

b) Make yourself familiar with the mentioned data set: Which information do you get about its structure by simply typing `us_empl_private` into the console?

c) Use the following graphics functions: `autoplot()`, `gg_season()`, `gg_subseries()`, `gg_lag()` and `ACF()` and explore features from `us_empl_private`.
Can you spot any seasonality, cyclicity and trend? What can you say about the seasonal patterns? Can you identify any unusual years and can you give an explanation for them? Refer to the different plots you have created to answer the questions.

3 Task 3

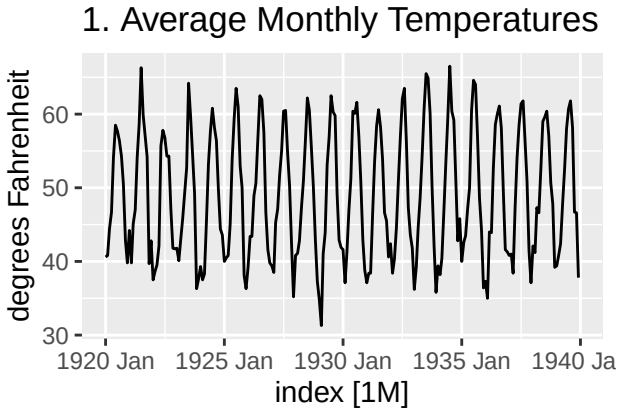
a) Use the following graphics functions: `autoplot()`, `gg_season()`, `gg_subseries()`, `gg_lag()` and `ACF()` and explore features from `Hare` from `tsibbledata::pelt` (O'Hara-Wild et al. 2022).

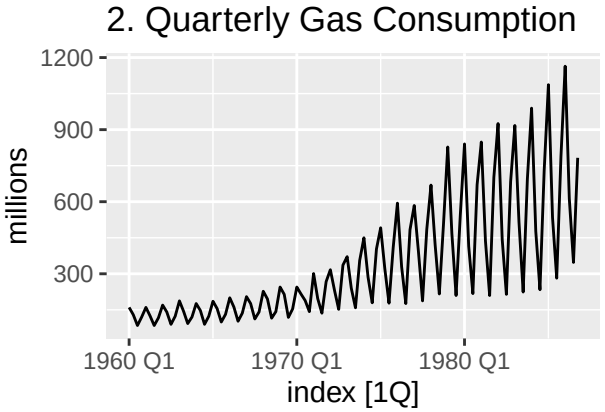
Can you spot any seasonality, cyclicity and trend? What can you say about the seasonal patterns? Can you identify any unusual years? Refer to the different plots you have created to answer the questions.

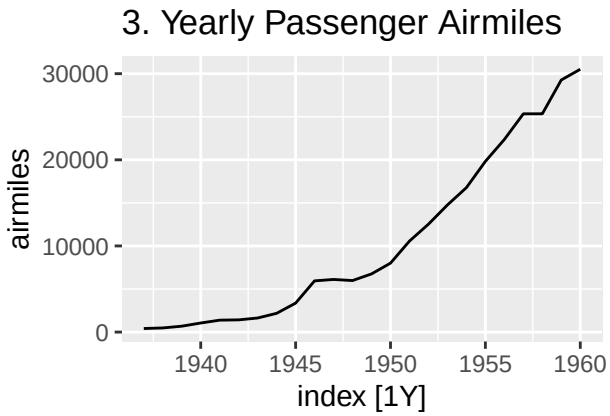
b) Again, plot the time series `Hare` from `pelt`. Add the time series `Lynx` from the same data set to the plot. What can you see when you compare both time series? Interpret it.

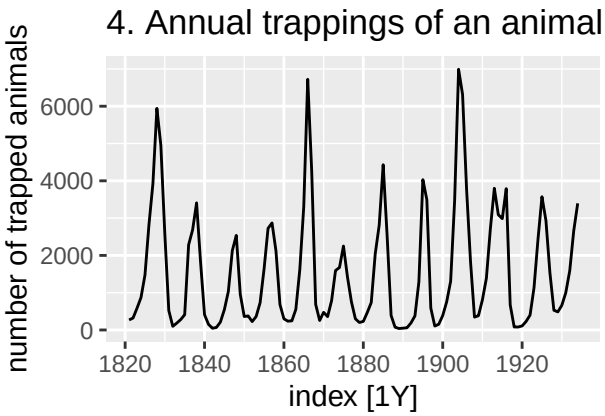
4 Task 4

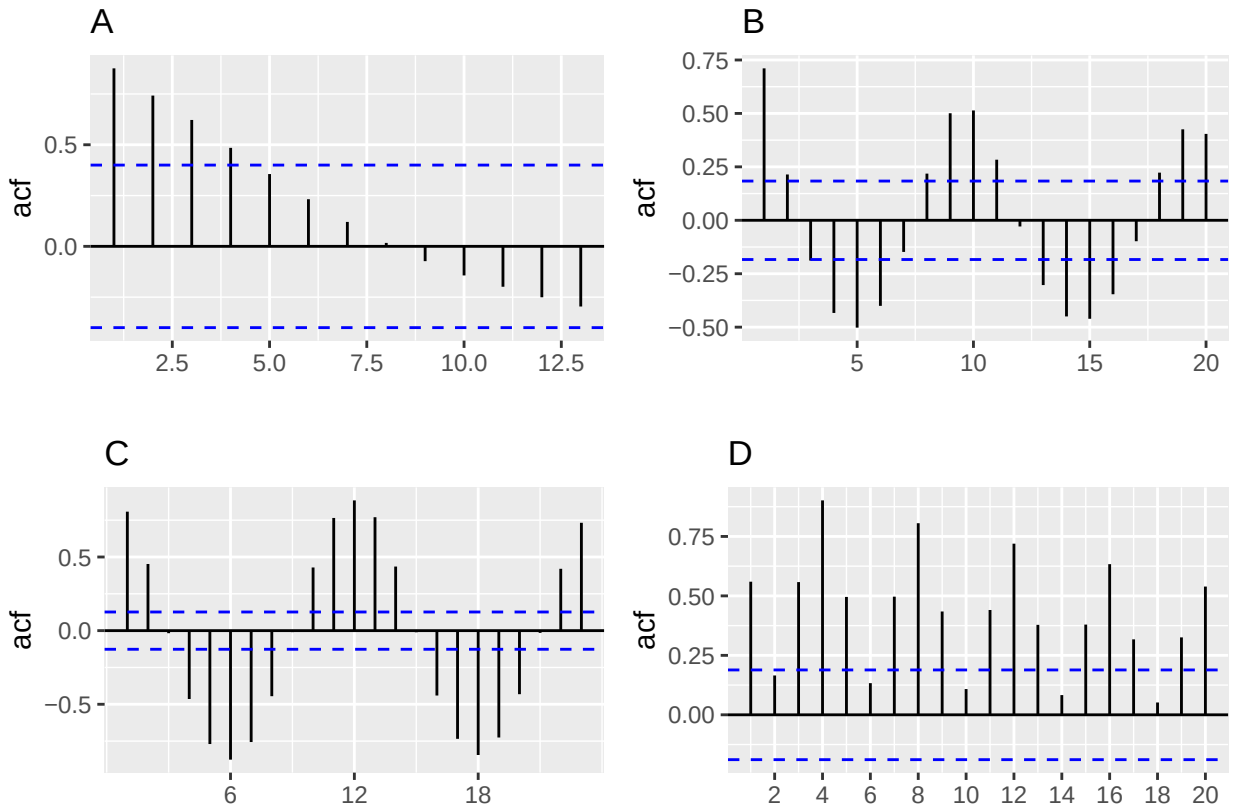
The following time plots and ACF plots correspond to four different time series. Match each time plot (1-4) with one of the ACF plots (A-D) and give an explanation for your decision!











5 Task 5

Write a function that computes the autocorrelation of a given time series as the function `feasts::ACF()` (O'Hara-Wild, Hyndman, and Wang 2021) does. Consider the argument `lag_max` of `feasts::ACF()` and include it in your self written function.
Check the functionality of your written function by applying it to the time series data `fpp3::canadian_gas` (R. Hyndman 2021).

Your computed autocorrelation coefficients should be identical to the following ones

```
canadian_gas_acf <- ACF(canadian_gas)
canadian_gas_acf
```

```
## # A tsibble: 27 x 2 [1M]
##         lag   acf
##    <cf_lag> <dbl>
##  1       1M 0.986
##  2       2M 0.974
##  3       3M 0.959
##  4       4M 0.945
##  5       5M 0.934
##  6       6M 0.925
##  7       7M 0.924
##  8       8M 0.924
##  9       9M 0.928
## 10      10M 0.931
## # i 17 more rows
```

and your corresponding ACF plot should look like this:

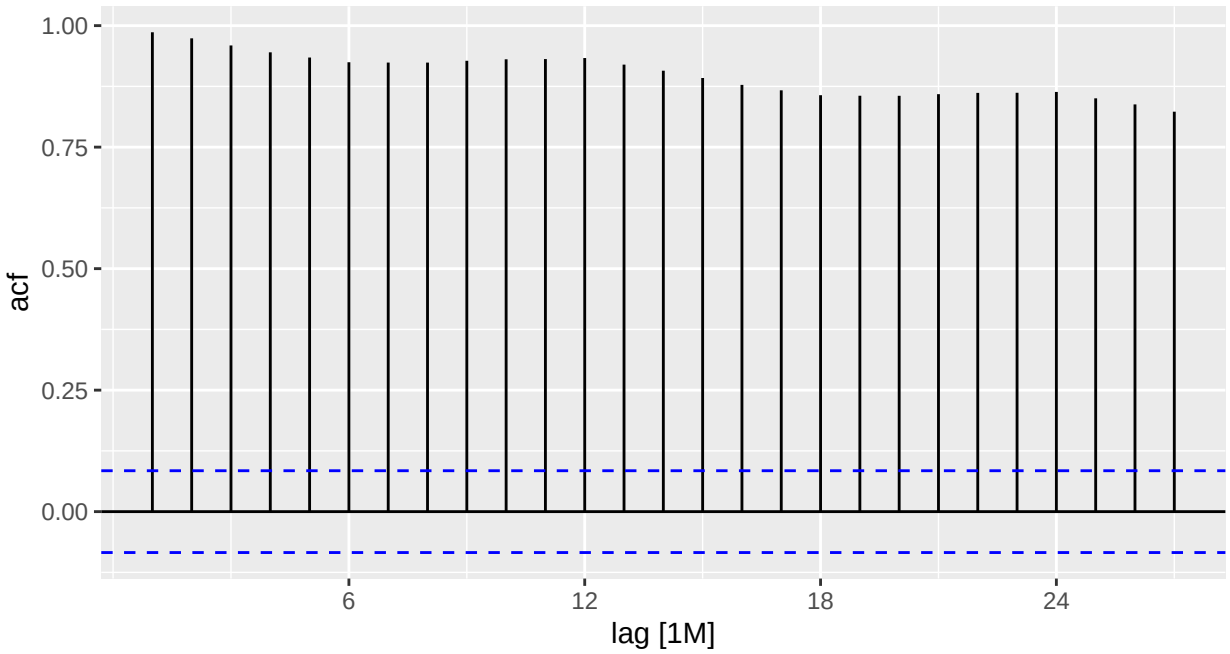


Figure 1: Autocorrelation plot based on the monthly canadian gas production (R. Hyndman 2021).

Interpret the resulting ACF plot (Figure 1)! *Hint*: Additionally, plot the time series `canadian_gas` for your interpretation.

6 Task 6

a) What is white noise? Mention its characteristics.

b) Describe and explain in your own words how you can generate a white noise series in R.
Then, generate a white noise series with 1,000 observations in R, convert it to a `tsibble` object (Wang, Cook, and Hyndman 2020) and plot it.

c) How can you control if you have really generated white noise? Describe it in your own words and do it!

d) What can you see in the ACF plot if you increase the number of observations of your generated white noise series, let's say from 1,000 to 10,000 and to 100,000? Give an explanation.

7 Task 7

Take again a look at the quarterly gas consumption time series (plot 2) from task 4.

a) Describe the time series.

b) What do you expect when you apply a Box-Cox transformation to this quarterly gas consumption data and why?

8 Task 8

Figure 2 shows the result of decomposing the number of persons in the civilian labour force in Australia each month from February 1978 to August 1995.

a) The recession of 1991/1992 is visible in the remainder component. What will you do if you want to consider the recession in the trend component?

b) Based on Figure 2, what can you say about the development of the number of persons in the civilian labour force in Australia each December from 1978 to 1994? Which component is relevant to answer the previous question?

c) How can you reconstruct the seasonally adjusted data of the time series from Figure 2? What do you have to do?

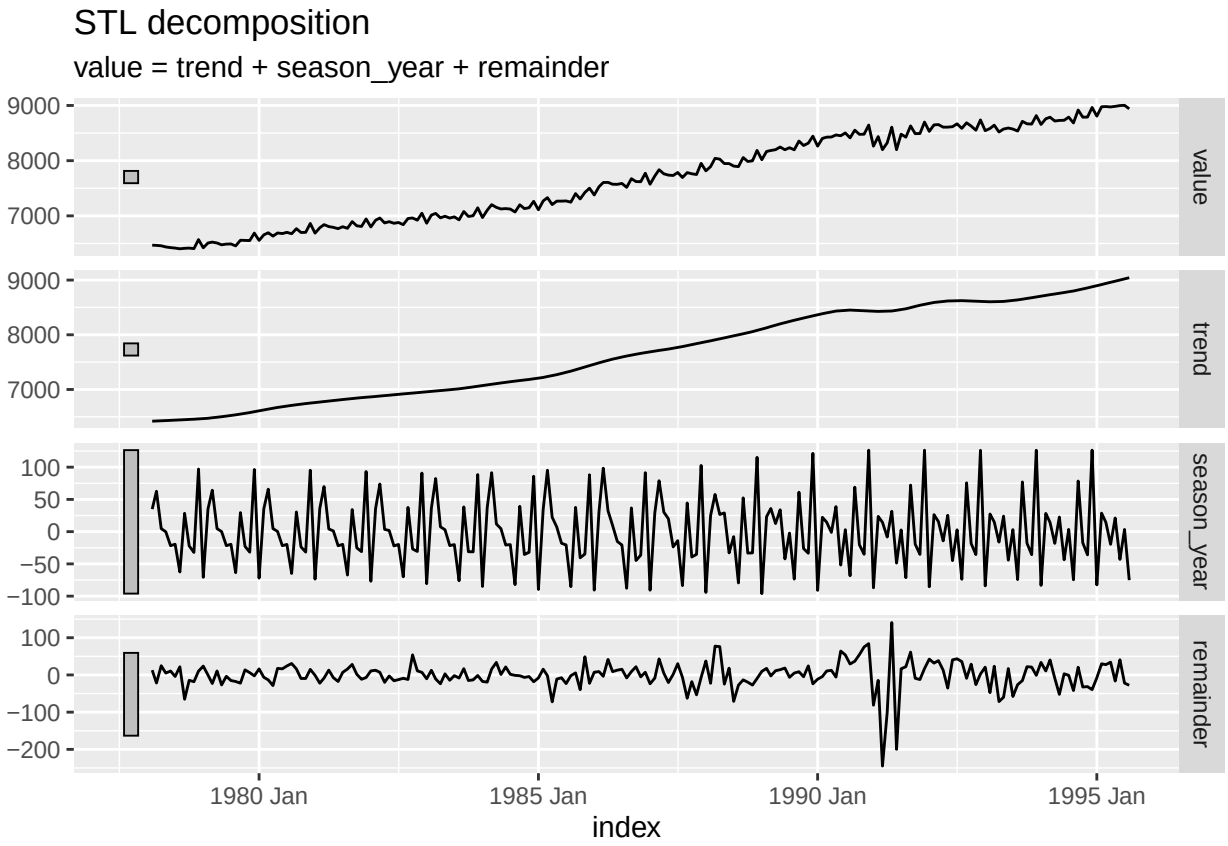


Figure 2: STL decomposition of the number of persons in the civilian labour force in Australia based on `fma::labour` (R. J. Hyndman 2020; Makridakis, Wheelwright, and Hyndman 1998).

9 Task 9

a) What does a 3x7 MA mean? Write down its equation!

b) Show that a 3x7 MA is equivalent to a 9-term weighted moving average with weights of 0.0476, 0.0952, 0.1429, 0.1429, 0.1429, 0.1429, 0.1429, 0.0952 and 0.0476.

c) Consider the time series `Spain` from the data set `tsibbledata::global_economy` (O'Hara-Wild et al. 2022) with `Exports` as dependent variable. Apply moving averages (MA) of different orders (at least two different orders) to this time series. Plot the time series and the different moving averages in one plot and answer the following questions:

c.1) Why do you apply a moving average to your data, i.e. what is the aim?

c.2) What happens when you increase the order m of the moving average and why does it happen?

References

Buchwitz, Benjamin. 2022. *Fhswf: Fhswf*.

Hyndman, Rob. 2021. *Fpp3: Data for "Forecasting: Principles and Practice" (3rd Edition)*. https://CRAN.R-project.org/package=fpp3.

Hyndman, Rob J. 2020. *Fma: Data Sets from "Forecasting: Methods and Applications" by Makridakis, Wheelwright & Hyndman (1998)*. http://pkg.robjhyndman.com/fma.

Kirchgässner, Gebhard, and Jürgen Wolters. 2007. *Introduction to Modern Time Series Analysis*. Berlin: Springer.

Makridakis, Spyros, Steve Wheelwright, and Rob J Hyndman. 1998. *Forecasting: Methods and Applications*. John Wiley & Sons.

O'Hara-Wild, Mitchell, Rob Hyndman, and Earo Wang. 2021. *Feasts: Feature Extraction and Statistics for Time Series*. https://CRAN.R-project.org/package=feasts.

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

Wang, Earo, Dianne Cook, and Rob J Hyndman. 2020. "A New Tidy Data Structure to Support Exploration and Modeling of Temporal Data." *Journal of Computational and Graphical Statistics* 29 (3): 466–78. https://doi.org/10.1080/10618600.2019.1695624.