EC 361–001

Problem Set 1

Prof. Santetti

Spring 2024

`INSTRUCTIONS`: Carefully read all problems. You must submit a single R script with your *first name(s)* (mine would be `marcio.R`). In case you submit your files with different names, you will lose 1 point.

You can find templates for your answer R script on `theSpring`, under the "Templates" module. Please consider using it.

I should be able to fully replicate your code to answer the questions, as well as fully understand your written interpretations to the proposed problems.

Avoid using unnecessary code in your submission files. It is totally fine to do other things by yourself that may help you better understand the data and the problems. However, for grading purposes, I am only interested in the commands and interpretations that actually answer the questions. You may keep a separate file for yourself with your additional explorations.

Recall that group work is *strongly encouraged* for going over this and all Problem Sets. For submission purposes, you may work *either* individually or in pairs.

`Assignment due Mar 1 (Fr), 12:20 PM`.
`Points Possible: 40`

- You have 2 weeks to complete this assignment. In accordance with our `course syllabus`, no late submissions will be accepted.
- Be honest. Don't cheat.
- As a Skidmore student, always recall your votes of academic integrity, and the **`Honor Code`** you have abided by:

> *"I hereby accept membership in the Skidmore College community and, with full realization of the responsibilities inherent in membership, do agree to adhere to honesty and integrity in all relationships, to be considerate of the rights of others, and to abide by the college regulations."*

Have fun!

Problem 1

The `ps1_data.csv` file (available on `theSpring`) contains three time series for the U.S. economy: `infrate`, the inflation rate measured by the Consumer Price Index (CPI); `unrate`, the civilian unemployment rate; and `fedfunds`, the federal funds rate. The sample period ranges from 01/1960 to 12/2023, in monthly frequency.

(a) Using the `read_csv()` function, import this data set into your workspace.

(b) Transform this data set into a time-series tibble (i.e., a `tsibble` object). Make sure to get the *frequency* correct.

(c) Generate a *time plot* of the U.S. unemployment rate.

(d) What relevant *features* do you observe when plotting the unemployment rate? When discussing features, focus on *trend*, *seasonal*, and *cyclical* components.

(e) Using the `filter_index()` and `autoplot()` functions, repeat part (c) only for the 01/2020—12/2023 window.

Problem 2

Install the `{gapminder}` package. It contains data on life expectancy, population, and GDP *per capita* for several countries. After you've installed and loaded the package, you just need to type `gapminder` in your workspace and you will be able to play around with the data. Then, answer the following questions:

(a) Transform the `gapminder` data set into a `tsibble` object. Since there are time series for different countries and continents, you will need to use the `key` argument.

(b) Using the `autoplot()` function, plot GDP *per capita* over time for the Chinese economy.

(c) Use the `mutate()` and `autoplot()` functions to plot together the *natural logarithms* of GDP *per capita* for the Japanese and Singaporean economies.

Problem 3

As seen in class, the `ACF()` function from the `{feasts}` package computes the autocorrelation coefficient for a time series. This coefficient is calculated as follows (considering a lag *k*):

$$r_k = \frac{c_k}{\text{Var(y)}}$$

where c_k is the autocovariance between a time series y_t and its k^{th} lag; and *Var(y)* is the variance of *y*.

For this problem, you will use the `retail_sales.csv` data set (available on `theSpring`). It contains quarterly data gathered by the U.S. Census Bureau on retail trade (including e-commerce). The length is 01/1992—12/2023.

Running the `ACF()` function for the first 5 lags of this series returns the following values:

```
retail_ts ▷
  ACF(lag_max = 5)
```

```
## # A tsibble: 5 x 2 [1M]
##        lag   acf
##   <cf_lag> <dbl>
## 1       1M 0.946
## 2       2M 0.924
## 3       3M 0.920
## 4       4M 0.913
## 5       5M 0.907
```

Using the `mutate()` and `summarize()` functions, *manually* compute the autocorrelation coefficients you see above for these data. *Hint 1*: Do not transform these data into a `tsibble` object; keep it as a `tibble`, otherwise the `summarize()` function does not work properly. *Hint 2*: You can use the `lag()` function to compute lags of any order you want. *Hint 3*: The `sum()` function computes (surprise, surprise), sums.

Problem 4

Still using the data set from **Problem 3**, answer the following questions:

(a) Transform the data set into a `tsibble` object.

(b) Using the `autoplot()` function, graph the data over time. What relevant *feature(s)* do you observe in this time series?

(c) Generate a *seasonal plot* for retail sales. How does it relate to the feature(s) you've observed in part (b)?

(d) Generate a *lag plot* for retail sales. How does it relate to the feature(s) you've observed in part (b)?

(e) Generate an *ACF plot* for retail sales. How does it relate to the feature(s) you've observed in part (b)?