EC 361–001

Problem Set 2

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 22 (Fr), 12:20 PM`.
`Points Possible: 40`

- You have 3 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

For this first problem, we will use once again the `ps1_data.csv` file (available on `theSpring`). It 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) Set your data set up as a `tsibble` object.

(b) Create a *training* set from your main data set, leaving out the last *12* observations.

(c) From the *training* set you've created in part (b), estimate 2 benchmark models for the inflation rate: *mean* and *naïve*. *Hint*: I recommend estimating the two models in the same `model()` statement.

(d) Produce a *12-month ahead* forecast.

(e) Using the `{patchwork}` package, plot the two forecasts you've generated in part (d) side by side. Use **95%** prediction intervals only. For these plots, limit your horizontal axis to *start only on 01/2018*. This way, you will better visualize your forecasts. Below, a way in which you can do it (using some imaginary names here):

```
my_forecast |>
  filter(.model == "whatever_method") |>
  autoplot(my_tsibble |>
             filter_index("Put your start date here" ~ .),
           level = 95) ## notice the use of the "pipe" operator
                       ## within the "autoplot()" function.
```

Problem 2

Still using your forecasts from **Problem 1**, answer the following questions:

(a) For both forecast methods, present the following *accuracy measures*: *mean absolute error*, *root mean squared error*, *mean absolute percentage error*, and *mean absolute scaled error*.

(b) From the measures you've shown in part (a), which model would you *select*? Explain.

(c) What is the *mean* value that each forecast method predicts for the "future" 12 months?

(d) Report *one* issue regarding the 95% prediction interval generated by your forecast using the *naïve* method. Explain it *intuitively*.

(e) What is the *main difference* between the two benchmark methods that you've used to forecast inflation? Explain it *intuitively*.

Problem 3

For this problem, you will use time series for U.S. **total business sales** (in millions of dollars). You can download the *not seasonally adjusted* series from `FRED`, searching for the `TOTBUSSMNSA` series code.

(a) Set your data set as a `tsibble` object. Make sure to have the correct *frequency*.

(b) Using the *classical decomposition method*, plot the original series and its 3 components: *trend-cycle*, *seasonal*, and *random*. Using the `{scales}` package, make sure to have an appropriate format to your *y-axis* values.

(c) From your answer to part (b), why does the graph for the *trend* component *starts after* and *ends before* the original series? Explain.

(d) Do you observe anything worth the attention in the *random* component you've generated? Explain.

(e) Repeat part (b), this time applying an *STL* decomposition method. Highlight *at least 2* main differences you observe between the two decomposition methods.

Problem 4

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

(a) These data are originally produced by the U.S. Census Bureau. `This page` contains information on their seasonal adjustment methods. From this document, apply the *same decomposition method* as the Census Bureau uses for *seasonal adjustment*.

(b) In the same plot, graph the original series and the seasonally adjusted component you've obtained in part (a). You can compare your results with the seasonally adjusted series available on FRED (its series code is `TOTBUSSMSA`). Make sure to include a legend distinguishing the two different series in your plot. *Hint*: Examples of graphing 2 (or more) series together with a legend have been shown several times in applied lectures.

(c) Store the `components` of the decomposition model you've applied in part (a) in a separate object, excluding the `.model` column. *Hint*: when you want to exclude a column from an object, you may run the following (using some imaginary names here):

```
your_new object <- some_other_object |>
  select(-column_name) ## the "-" sign facilitates our lives.
```

After doing that, produce a 24-month ahead forecast for the **seasonally adjusted** series only, using the *naïve* method.

(d) What is the (average) total sales value predicted by your part (c)'s model?

(e) Run a *Box-Pierce* and a *Ljung-Box* test on the *residuals* from your forecasting model. Do you have *evidence* in favor of these residuals being *white noise?* Explain your reasoning.