STAT 153 Project Checkpoint 2

Assigned: March 3, 2021

Due: March 10, 2021, by 11:59pm
(5% off per rounded-up hour late, like HW)

For the second project checkpoint, you will need to write most of the first 2.5 sections of your report (roughly the first half). It does not need to be perfect by any means, but it should all be attempted! These sections are:

1. First is the executive summary, which will not yet state what your final model and forecasts are but should state the problem.
2. Second is exploratory data analysis (EDA), which needs to contain a time series plot of your chosen dataset (either "stocks" or "covid") that you already made for checkpoint 1. To this you should add your own description of the relevant features of your data, particularly pointing out what features are preventing it from being stable/stationary (i.e. trend, seasonality, heteroscedasticity).
3. The third and largest section of your report is the description of the chosen models. For this checkpoint you should do the first main step here: build two signal models. Your signal models can be anything we have discussed or even something beyond the scope of this course if you want. Usually, students use 1) a parametric model and 2) a differencing approach, but I only state this to illustrate how you might think of it. You are welcome to use smoothers too of course.

Please **review the project documents on bCourses** for further details! And remember, though you will not be graded on formatting for this particular checkpoint, a professional-looking document is the end goal (as I'm sure your end goal after school is to be professional, wherever your education takes you!).

This checkpoint is worth **5 points** total, and it's simple to get full points because it is completion graded! The rubric/breakdown of these five points is:

1. Includes everything from Checkpoint 1 (PDF format, professional heading, plot of the dataset; if you did CP1, this is a free point!)
2. Executive summary contains appropriate text.
3. EDA section contains appropriate text.
4. Signal model 1 is described in words or stated mathematically, and has a reasonably-stationary residual plot or differences plot.
5. Signal model 2 is described in words or stated mathematically, and has a reasonably-stationary residual plot or differences plot.

   (Note on your signal models: they don't need to be perfect or what you'll use on the final submission. But, they should not have obvious non-stationary patterns of trend, seasonality, or heteroscedasticity. We know how to approach these issues! If you cannot solve these issues, these challenge(s) should be discussed in the paper and efforts from trying multiple options should be described clearly.).

If you have any questions, please ask!