Instructions

(i) Do either Question 1 (specific) or Question 2 (open-ended). You should typeset a report of your findings (short paper), following the guidelines and template at the top of the page: `https://www.math.ttu.edu/ atrindad/tsdata/`.

(ii) Collaboration is allowed, but only in general terms, e.g., networking for troubleshooting of code is ok, but copying exactly what someone else is doing is not. You should write-up your results as a set of slides to be presented in class, and plan to take up a maximum of 25 minutes. Your grade will be based on scientific accuracy, completeness, and presentation.

(iii) Data file **Boeing.csv** can be found on:

`http://www.math.ttu.edu/~atrindad/tsdata/Boeing.csv`

1. The file **Boeing.csv** contains information on Boeing stock trades for the period of 5 days spanning December 1 to 5 of 2008. The stock is traded on the NYSE which is open continuously from 9:30 to 16:00 on weekdays. Read this in as follows:

```
> Bt=read.csv("Boeing.csv")
> head(Bt, 5)
   year month day hour minute second price volume                time
1  2008    12   1    9     30      2 41.64   3441 2008-12-01 09:30:02
2  2008    12   1    9     30      2 41.64   3441 2008-12-01 09:30:02
3  2008    12   1    9     30     12 41.43    100 2008-12-01 09:30:12
4  2008    12   1    9     30     16 41.82    100 2008-12-01 09:30:16
5  2008    12   1    9     30     16 41.83    100 2008-12-01 09:30:16
```

Each line (row) of the file corresponds to a trade/transaction, and there are 224,326 of them. We can see that in going from trade #1 to trade #2 there was no price or volume change, and the duration was 0 seconds because they occurred at the same time (09:30:02). In going from trade #2 to trade #3 there was both a price and volume change (100 shares were traded at $41.43 each), and the duration was 10 seconds (the time difference between 09:30:02 and 09:30:12).

The goal is analyze the **durations** data, and ultimately fit the autoregressive conditional duration (ACD) models, originally proposed by Engle & Russell (1998), using the capabilities of the R package `ACDm`. The first step is to extract the durations with the command:

```
> library(ACDm)
> dur=computeDurations(Bt, open = "09:30:00", close = "16:00:00",
                       rm0dur = TRUE, type = "trade")
> head(dur, 5)
The 224326 transactions resulted in 54795 durations
                  time    price volume Ntrans durations
1 2008-12-01 09:30:02 41.64000   6882      2         2
2 2008-12-01 09:30:12 41.43000    100      1        10
3 2008-12-01 09:30:16 41.91116   2016     11         4
4 2008-12-01 09:30:17 41.97000   1000      3         1
5 2008-12-01 09:30:27 41.50500    200      2        10
```

Use this code exactly as shown, the open and close times are specific to NYSE, and we only want to extract positive durations (setting `rm0dur = TRUE` removes the zero durations).

In your analysis you should consider the following:

(a) Plot the data in meaningful ways and describe its salient features. Is there a diurnal cycle/trend?

(b) If there is a cycle and/or other systematic trends, you will need to model it/them and extract the residuals before proceeding with ACD modeling. One possibility is to construct your own trend functions as on p. 253 of the AFTS book. Another option is to use the `diurnalAdj` function in `ACDm` to perform nonparametric smoothing. If you use the latter, note that the `aggregate` option is day specific, e.g., if "all" then all days are smoothed in the same way, if "none" then each day is smoothed differently, etc.

(c) Fit ACD models using the function `acdFit`. Explore the range of available models (see help file), and decide on what you believe to be the "best" model. Note that you have several "plot" functions and other diagnostics to help you answer this question.

(d) Is it possible to do foreasting with your model? Explore this and attempt to forecast the next 3 durations. (You may need to consult the literature...)

2. Propose a specific project that involves the analysis of some dataset with research questions in mind. Especially encouraged is high-frequency data. You should include the following goals:

   - Where will you get the data?
   - What questions will you seek to answer?
   - What will you analyze/what models will you use?

   [Note: one source of high-frequency data is: `https://lobsterdata.com/`]

References

[1] Engle, R.F. and Russell, J.R. (1998), "Autoregressive conditional duration: a new model for irregularly spaced transaction data", *Econometrica*, 66, 1127–1162.