Homework #3
Due on Wednesday, April 10, 2024
(submit by uploading pdf to Canvas)

Organize your homework following homework_guidelines.pdf.

The 4 data sets BOND, GEYSER, REPAIR and FAKE each contain a single time series. (These time series are of varying lengths.) Choose a reasonable ARIMA model for each of these series. Note that some of the series may require a transformation.

For each time series, submit the following <u>in this order</u>:

- A summary sheet stating your choice of model and describing in detail your reasons for choosing this model. (This should be the first page of each problem solution.)

  The description of your model should include the transformation you used (if you used any), the values of p, d, q, and whether or not your model contains a constant C. (Write `NOCONSTANT` if you drop the constant.) A sample description is the following:

  For the series $Y = \sqrt{X}$, I chose an ARIMA(3,2,1) model without a constant (`NOCONSTANT`) .

  In each problem, in your SAS code and homework please refer to the raw time series (without any transformation or differencing) as X.

- All the usual output produced by running PROC ARIMA on your chosen model.

  This will include the ACF, PACF, IACF and time series plot for the (possibly transformed) series differenced d times (where d is your final choice for the degree of differencing) and all the estimation results and residual diagnostics for your model. You should also include the time series plot of the residuals, which will be produced by the `PLOTS=ALL` option in `PROC ARIMA`. (See `hw3_template.sas`.)

- The plot of residuals versus predicted values for your chosen model.

- Any other output you explicitly refer to in your summary statement.

All output should be annotated. Important parts of the output should be indicated and you should explain what conclusions you draw from this indicated output. Use complete sentences please.

Each series is a separate problem:

Problem 1: BOND

Problem 2: GEYSER

Problem 3: REPAIR

Problem 4: FAKE

The data sets BOND, GEYSER, REPAIR, and FAKE are available in the hw3 folder in mordor or (if you use SAS Studio in the cloud) you can access them in the course folder by following the example in `hw3_template.sas`.

Remarks on HW3 (also given in the hw3 folder in mordor)

AIC and SBC: You should only use AIC or SBC to compare models which involve the same transformation (if one is used) and the same degree of differencing. You will lose some credit if you use AIC or SBC to compare models involving different transformations or degrees of differencing. Also note: if you want to compare models using AIC or SBC, it is best to use METHOD=ML in the ESTIMATE statement. The other estimation methods only compute an approximation to the AIC or SBC.

Warning Messages: If you try a model and there are warning messages in the SAS output indicating failure to converge or some such problem, you should try a different model. You will lose some credit if you choose a model which produces such warning messages (if there are other models which look reasonable which do NOT produce such messages).

Non-significant terms: We usually use P-value $< .05$ as our standard for statistical significance. Terms which are NOT statistically significant are usually dropped from the model. Sometimes terms which are not quite significant (the P-value is close to but not less than .05) are retained in the model if there is some good reason to, e.g., if they lower the AIC or greatly improve the residual ACF. In your homework, if you retain any non-significant terms, you should give a good reason for doing so. If you neglect to state a good reason, you will lose some credit.

Seasonal Effects: BOND and REPAIR are monthly time series and contain likely seasonal effects which show up as moderately large residual autocorrelations at lag 12 (and perhaps lag 24 also). However, in this homework you are NOT supposed to use seasonal models, and so we just have to live with these problems; they will probably exist in all non-seasonal models that are tried. So don't try hard to fix any residual correlations that show up around lags 12 or 24.