Homework #4
Due on Friday, April 26, 2024
(submit by uploading pdf to Canvas)

Organize your homework following homework_guidelines.pdf.

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

Problems 1(a), 2(a), and 3(a) require you to find a "reasonable model" for a time series. In 1(a), 2(a), and 3(a), to describe and document your model selection process, please submit the following **in this order**:

- Your choice of model.

  The description of your model should include the transformation you used (if you used any), the values of p, d, q, P, D, Q, and whether or not your model contains a constant C. Write `NOCONSTANT` if you drop the constant. In problem 3(a) you must also specify which sine and cosine terms you included in your trend function.

  A sample description is the following:

  > For the series $Y = \sqrt{X}$, I chose an ARIMA$(2,1,3)(1,0,2)_{12}$ model without a constant (`NOCONSTANT`) .

- A detailed description of your reasons for choosing this model.
- All the usual output produced by running PROC ARIMA on your chosen 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 the detailed description of the reasons for choosing your model given earlier.

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.

Problem 1: We revisit the REPAIR series from HW3. This series is monthly data and contains likely seasonal effects which we did not attempt to model in HW3. However, we did consider transformation and differencing in HW3, and to avoid repeating this work, for the REPAIR series everyone should use a log transformation (i.e., model $y_t = \log(x_t)$) and the differencing `d=1, D=0`. So, for this series you need only choose p, q, P, Q and whether or not your model contains a constant C.

(a) Find a reasonable model for this series which includes a seasonal term. (Throughout this problem, use `NLAG=24` when running `PROC ARIMA` so that you get to see the first two seasonal lags.)

(b) Compare the model you found in part (a) to the same model without the seasonal term. How does adding the seasonal term affect the residual diagnostics? The AIC/SBC?

(c) Consider two versions of the seasonal model you found in part (a), one including the constant (C) and one without the constant. Use each model to forecast 36 months into the future. How do the forecasts and the 95% confidence intervals differ for these two versions of the model? (In this problem, leave all the forecasts and confidence intervals on the transformed scale; do not back-transform.)

Problem 2: The time series BUS.TXT is monthly data giving the average number of bus passengers per weekday.

(a) Find a reasonable seasonal ARIMA model for this time series.

(b) Use this model to give forecasts for 36 months beyond the end of the series. If you use a transformation, make sure to give forecasts and confidence intervals on the original scale. (Unlike in the previous problem, here I am asking you to do the back-transformation.)

Problem 3: The time series EUREKA.TXT is monthly data giving average temperatures.

(a) Find a reasonable model of the form (trend)+(stationary ARMA process) for this series. The "trend" should be a periodic function (repeating every 12 months) consisting of a sum of sines and cosines. The "stationary ARMA process" could be either non-seasonal or seasonal; pick whatever gives the best fit. (This model does **not** use any transformation or differencing.)

(b) Use this model to give forecasts for 36 months beyond the end of the series. Describe the long-run behavior of the forecasts and the associated standard errors and confidence interval widths.

Remarks on HW4 (also given in the HW4 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.