Name: Your name here
Due: 2024/11/20

Day 20 - Lab: ARMA models

Introduction

In this assignment, we will get some more practice with analyzing real data using time series models that incorporate ARMA(p, q) processes. In this lab, we will return to the electricity data that were introduced when decomposing time series at the start of the courses. As a reminder, this data set describes the monthly supply of electricity (in millions of kWh) in Australia over the period of January 1958 to December of 1990, according to the Australian Bureau of Statistics. The code below reads in the series.

```
cbe <- read_delim("cbe.dat")
elec <- dplyr::select(cbe, elec)
```

---

1. [2 pt] Create a data frame that includes the electricity measurement, the month, a scaled index for time (to have mean 0 and standard deviation 1), and the scaled time index squared.
2. [2 pt] Plot electricity over time and describe the series in terms of trend and seasonality.

```
```

3. [4 pt] Fit a regression model of the form `elect ~ scaled_t + scaled_t2 + month` and assess the assumptions for the fitted model.

```
```

4. [2 pt] Fit the same model again, this time using `log(elec)` as the response. Which assumptions are still violated?

```
```

5. [2 pt] Ignore any violations of the assumptions for now (except for independence) and use the log-transformed structure for all remaining questions. Use `arima` to fit the regression model with an AR(1) correlation structure. Assess the residual serial correlation.

```
```

6. [2 pt] Use `arima` to fit the regression model with an ARMA(1, 1) structure. Assess the residual serial correlation.

```
```

7. [4 pt] Use the `auto.arima` function in the `forecast` package to pick the best non-seasonal ARMA model (set `max.d, max.D, max.P, max.Q` all equal to 0) for this regression model and print off the fit. Assess the residual serial correlation. You will need to set `allowmean = F` and `allowdrift = F` in the function call, since both the mean and the drift are included in our regression model (the y-intercept and the slope with time).

```
```

8. [4 pt] Allow `auto.arima` to choose any form for the errors in the regression model by no longer setting `max.d, max.D, max.P, max.Q` equal to 0 (might take a second or two). You will again need to turn off the drift and mean parameters. Print off the fit and assess the residual correlation.

```
```

9. [4 pt] Finally, omit the regression model all together and allow `auto.arima` to select the best model. Print off the model and assess the residual correlation.

```
```

10. [2 pt] We will learn about the model that `auto.arima` selected in question 8 and 9 next week. For now, comment on the advantages and disadvantages of the purely stochastic approach implemented by `auto.arima` with no `xreg` relative to including predictor variables in a regression model.