ST5209/X Assignment 5

Due 22 Apr, 11.59pm

Set up

1. Make sure you have the following installed on your system: $\LaTeX$, R4.2.2+, RStudio 2023.12+, and Quarto 1.3.450+.
2. Pull changes from the course repo.
3. Create a separate folder in the root directory of the repo, label it with your name, e.g. `yanshuo-assignments`
4. Copy the assignment1.qmd file over to this directory.
5. Modify the duplicated document with your solutions, writing all R code as code chunks.
6. When running code, make sure your working directory is set to be the folder with your assignment .qmd file, e.g. `yanshuo-assignments`. This is to ensure that all file paths are valid.[1]

Submission

1. Render the document to get a .pdf printout.
2. Submit both the .qmd and .pdf files to Canvas.

1. Modeling with ARIMA (Q9.7 in Hyndman & Athanasopoulos)

Consider `aus_airpassengers`, the total number of passengers (in millions) from Australian air carriers for the period 1970-2011.

a. Use `ARIMA()` to find an appropriate ARIMA model. What model was selected. Check that the residuals look like white noise. Plot forecasts for the next 10 periods.

b. Write the model in terms of the backshift operator.

c. Plot forecasts from an ARIMA(0,1,0) model with drift and compare these to part a.

[1] You may view and set the working directory using `getwd()` and `setwd()`.

d. Plot forecasts from an ARIMA(2,1,2) model with drift and compare these to part b. Remove the constant and see what happens.

e. Plot forecasts from an ARIMA(0,2,1) model with a constant. What happens?

2. Recursive forecasting for ARMA(1, 1)

Consider the $ARMA(1,1)$ model

$$X_t - 0.5X_{t-1} = W_t + 0.5W_{t-1}.$$

In this question, we will investigate recursive forecasting. The following code snippet generates a sequence of length $n = 50$ drawn from the above model.

```r
set.seed(5209)
n <- 50
wn <- rnorm(n)
xt <- arima.sim(model = list(ar = 0.5, ma = 0.5), innov = wn, n = n)
```

a. Fill in the following code snippet using equation (11.14) to generate a sequence `wn_hat`.

```r
wn_hat <- rep(0, n)
wn_hat[[1]] <- xt[[1]]
for (i in 2:n) {
  # FILL IN
}
```

b. Make a time plot of the log absolute difference between `wn` and `wn_hat`.

c. What consequence does this have for truncated forecasts?

d. Compute the truncated forecast for X_{53}.

3. Seasonal ARIMA

a. Load `diabetes.rds` from the directory `_data/cleaned`.

b. Perform the following transformation of the column `Cost`: Apply a log transform followed by a seasonal difference. Label the resulting time series `Y`.

c. Apply the KPSS test to `Y`. What is its p-value? What can you conclude about `Y`?

d. Make a time plot of `log(Cost)`. Why does the trend disappear when we consider `Y`?

e. Fit an ARIMA model to `log(Cost)` and report the order of the fitted model.

f. How many fitted parameters are there in the model?

4. Model selection

a. What are the null hypothesis and assumptions of the ADF test?

b. Is it possible for both the ADF and KPSS test applied to a dataset to have large p-values? Explain why or why not.

c. What are the AIC and AICc penalties for the model fitted in Q3?

d. Fit an exponential smoothing model of your choice to `diabetes.rds`. Use `glance()` to view the log likelihood and AICc values of both this model and the ARIMA model from Q3.

e. Can we say which method is a better fit to the data by comparing their log likelihood or AICc? Explain why or why not.

5. ACF, PACF, and BLPs

Let (X_t) be a mean zero stationary process with the following autocovariance values:

$$\gamma_X(0) = 2, \gamma_X(1) = 1.4, \gamma_X(2) = 0.6, \gamma_X(3) = 0.4, \gamma_X(4) = 0.2.$$

a. Can (X_t) be an MA(2) process? Explain why or why not.

b. Can (X_t) be an AR(1) process? Explain why or why not.

c. What is the best linear predictor $\hat{X}_4$ for X_4 given only $X_3 = 2$?

d. Using the notation in part c), what is the variance of $X_4 - \hat{X}_4$?

e. What is the best linear predictor $\hat{X}_4$ for X_4 given only $X_2 = 2$?

f. f. Using the notation in part e), what is the variance of $X_4 - \hat{X}_4$?

g. Let α_X denote the partial autocorrelation function of (X_t). What is $\alpha_X(1)$?

h. What is $\alpha_X(3)$?