Homework 4

Note: The due date is February 29 (Thursday). The goal of this homework consists of two. First, I expect you to extend your knowledge of the ARMA class to the ARIMA class from a modeling perspective. Second, I expect you to practice the procedure of the unit-root tests for a given time series. Any questions about R/RStudio, email the teaching assistant. All problems below will carry equal weight.

Problem 1. Recall that the difference operator Δ is defined by $\Delta X_t = (I - B)X_t = X_t - X_{t-1}$. Likewise, $\Delta^k X_t = (I - B)^k X_t$, $k \geq 1$ and $\Delta_s X_t = (I - B^s)X_t = X_t - X_{t-s}$, $s \geq 1$. Do the following.

(a) Suppose $X_t = \beta_1 + \beta_2 t + \ldots + \beta_k t^r + Y_t$ with $\beta_k \neq 0$, where $\{Y_t\}$ is a stationary process. Show that $\Delta^k Y_t$ is stationary for $k \geq 1$ and $\Delta^k X_t$ is stationary for $k \geq r$. If needed, you can denote the ACVFs of $\{X_t\}$ and $\{Y_t\}$ as $\gamma_X(h)$ and $\gamma_Y(h)$, respectively.

(b) Let $X_t = a + bt + S_t + Y_t$, where a, b are constants, S_t is a seasonal component with period 12, and $\{Y_t\}$ is a stationary process with mean zero. Denote the ACVF of $\{Y_t\}$ as $\gamma_Y(h)$. Show that $\Delta_{12} X_t$ is stationary; Express its ACVF in terms of $\gamma_Y(h)$.

Problem 2. Consider the time series data `gnp` from the R package `astsa`. Do the following.

(a) Take a suitable preliminary transformation of the series, and produce its time plot; In the following parts, work with the transformed series.

(b) Leave out the last 5 observations. Denote these samples as test data. The remaining observations are your training data. Fit a quadratic trend to the series using regression with the training data; Produce a time plot and a correlogram of the residuals obtained after removing the trend from the series;

(c) Fit an ARMA(p, q) model to the residual of the regression with an order p, q determined by an information criterion; Include the output; Produce the sample ACF and PACF of the residuals of the ARMA(p, q) model; Check the assumptions of white noise and normality for the residuals;

(d) Forecast the transformed time series for 5 steps into the future; Compute the mean squared forecast error (MSFE) by using the test data; Provide the outputs;

Are we forecasting trainigndata here?

or residuals?

If residuals then what does it make sense that we fit a model to the residuals and use it to model the original series

(e) Now, fit an ARIMA(p, d, q) model to the training data; Forecast the transformed time series for 5 steps into the future by using this model; Compute the MSFE again; Which model is preferred in terms of minimal MSFE? Include the output;

Problem 3. Do the following.

(a) Use `set.seed(99)` to generate the following random walk with drift:

include.drift = TRUE
allowdrift = TRUE

$$X_t = -0.2t + 0.8\sum_{s=1}^{t} Z_s, \quad t = 1, \dots, 100,$$

is q=100 or less here?

where $\{Z_t\}$ is IID standard normal; Include a time plot of the series and sample ACF and PACF of the series.

allowmean ?

Should we inclide?

Are we using ARMA(pq) or ARMA(pdq)?

(b) Use the function `auto.arima` with suitable inputs to recover this model. That is, can you find an outcome of the model indicating the random walk with drift?

(c) Go through the testing procedure for unit roots with significance level $\alpha = 0.05$ (for all steps). Indicate the conclusion at each step of the procedure; Check if the testing result corresponds to the used model;

Problem 4. Do the following real data applications.

(a) Consider the time series `Raotbl3$lc` of real consumption expenditure from the United Kingdom starting in 1966:4 until 1991:2 in the R package `urca`. Produce a time plot of the series; Go through the testing procedure for unit roots discussed in class taking $k = 3$ for the number of lagged series differences to include in the regression; Indicate the conclusion at each step of the procedure.

(b) Repeat the testing procedure with the series in (a) with a smaller lag, $k = 2$; Check if a different choice of lag affects the conclusion;

(c) Go through the testing procedure with the transformed `gnp` used in Problem 1. Here, use the entire samples. Use the lag p determined for the ARMA model. If your ARMA model contains the MA part, use `selectlags`, contained to `ur.df` function; Check if the conclusion of this problem corresponds to the preference of the model (i.e. trend stationary or non-stationary) in Problem 1. (e);

Do we need to fit anther model here or use the one we trained in 2c?