STAT 443 – Winter 2015 – Assignment 3

due Thursday March 12 at the beginning of class

You may work in pairs if you choose; both names and ID numbers should appear on it, and both will receive the same mark. (No extra credit will be given for working alone.)

For any parts involving R, you should hand in the R code and output, as well as your interpretations of the output. You will NOT receive marks for uncommented R code or output.

1. A time series $X_1, ..., X_{100}$ has sample mean $\bar{X} = 46.93$ and sample autocovariances $\widehat{\gamma(0)} = 1382.2$, $\widehat{\gamma(1)} = 1114.4$, $\widehat{\gamma(2)} = 591.73$, and $\widehat{\gamma(3)} = 96.216$.

   (a) Use these values to find the Yule-Walker estimates of ϕ_1, ϕ_2, and σ^2 in the $AR(2)$ model $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t, \{Zt\} \sim WN(0, \sigma^2)$ for the mean-corrected series $Y_t = X_t - 46.93, t = 1, ..., 100$.

   (b) Assuming that the data really are a realization of an $AR(2)$ process, find 95% confidence intervals for ϕ_1 and ϕ_2.

   (c) Compute the sample partial autocorrelations $\widehat{\alpha(1)} = \widehat{\underline{\phi_1}_1}$, $\widehat{\alpha(2)} = \widehat{\underline{\phi_2}_2}$, and $\widehat{\alpha(3)} = \widehat{\underline{\phi_3}_3}$ of the series Y_t. Is the value of $\widehat{\alpha(3)}$ consistent with the hypothesis that the data are generated by an $AR(2)$ process?

2. The average daily temperature of a small town in England over a period of about 4 months is provided in the file *temp.txt*. Use the following commands to load the data into R.

   ```
   temp <- read.table("temp.txt")
   temp <- temp[,1]
   ```

   In the context of prediction, the average sum of squares of one-step forecast error is given by $PMSE = \frac{\sum_{t=1}^{n-1}(X_{t+1} - \widehat{X}_{t+1})^2}{n-p}$, where p is the number of parameters of the model.

   (a) Using simple exponential smoothing, the forecast of X_{n+1} based on $X_1, ..., X_n$ is $\widehat{X}_{n+1} = \widehat{\beta}_0(n) = \alpha X_n + (1-\alpha)\widehat{\beta}_0(n-1)$ where $\widehat{\beta}_0(0) = X_1$. Write an R function to calculate the sum of squares of one-step forecast error based on each value of $\alpha \in \{0.01, 0.02, ..., 0.99\}$.

   (b) Plot the values of $PMSE$ versus α from part(a) and determine the best value for α.

   (c) Use the Holt-Winters built-in function in R to perform simple exponential smoothing, and compare the optimal α from the HW function to your optimal α in part (b).

   (d) Use the Holt-Winters function to perform double exponential smoothing and calculate the PMSE statistic. Comparing this to the model in (c), which model provides a better prediction: simple exponential smoothing or double exponential smoothing?

3. The time series built into R USAccDeaths contains the monthly number of accidental deaths in the US for several years. Insurance companies would be interested in predicting the future number of accidental deaths, as usually the payout for insurance is higher if death is accidental. You are a statistical consultant and you are going to consider several models for prediction and then report to the insurance company.

   (a) Split off the last year of data (12 observations) to use as a testing set for your data. You will build all your models using the traning set of the first 60 observations only. Plot the training set and comment on what the data shows.

(b) First you consider a classical decomposition model: $Y_t = T_t + S_t + \epsilon_t$ where T_t is the trend, S_t is the seasonal effect, and ϵ_t are the errors.

  i. Fit a regression model including linear and quadratic functions of time, and indicator variables for the month.
  ii. Check the model adequacy using residual plots.
  iii. Predict the last 12 observations in the original series (61 to 72) with 95% prediction intervals. Does the testing set fall in the intervals?
  iv. Compute the Prediction Error Sum of Squares (PRESS) for the testing set.

(c) Next you consider using Holt-Winters smoothing but you are not sure whether the seasonal variability is constant.

  i. Fit both seasonal Holt-Winters models with additive seasonality and with multiplicative seasonality.
  ii. Check the adequacy of both models using residual plots.
  iii. Use both models to predict the last 12 values with 95% prediction intervals. Does the testing set fall in the intervals?
  iv. Calculate the PRESS for both models.

(d) Finally you consider a SARIMA (seasonal ARIMA) model.

  i. Determine the amount of seasonal differencing (at lag 12) D and ordinary differencing d to make the data stationary.
  ii. Using the sacf and spacf of the differenced series, determine the appropriate values of p, q, P, and$Q$ for a seasonal $ARMA$ model for the differenced series. Fit it and check it using residual plots.
  iii. Fit the $SARIMA(p, d, q) \times (P, D, Q)_{12}$ model to the original data using the arima function: model <- arima(series,order=c(p,d,q),seasonal=list(order=c(P,D,Q),period=12),method="ML")
  iv. Predict the last 12 values with 95% prediction intervals. Does the testing set fall in the intervals?
  v. Calculate the PRESS.

(e) Which is the final model you propose? Present your conclusions to the insurance company in the form of a short paragraph and a plot of your chosen model's predictions for the **next** 12 months (beyond the end of the testing set.)