Statistics 153 (Introduction to Time Series) Homework 6

Due FRIDAY, April 30, 2021, by 11:59pm

1. **Model Diagnostics and Selection**
   Consider the time-series dataset contained in the `gas_data.csv` file on bCourses under Files > Datasets. This file contains the Australian monthly gas production for years 1970-1995. (This is a subset of the dataset found in the `forecast` package in R.)

   (a) Fit a SARIMA model to this data. You may need to use a variance stabilizing transform, but you will certainly need to take difference and/or seasonally difference the data. The choices of p/d/q/P/D/Q/S are yours to make. The model need not be perfect, but you should provide the sarima() diagnostic plots and describe why the plots show that your model fits *reasonably* well (again, reasonable-but-not-perfect is fine). *(3 points)*

   (b) To make things more interesting, suppose you don't believe in ARIMA modeling and consider doing a curve fitting with a third-degree polynomial plus nonparametric seasonal components instead. In particular, consider the following model:

   $$\text{Model 2: } V_t = \beta_0 + \beta_1 t + ... + \beta_3 t^3 + \beta_4 I(t \text{ is January}) + ... + \beta_{14} I(t \text{ is November}) + w_t \tag{1}$$

   where (w_t) is iid $N(0, \sigma^2)$. To demonstrate model fit, show the residual plot, ACF, and PACF. Note that it's okay to check the ACF at this point as we are assuming the residuals are stationary white (Gaussian) noise... but is that a good assumption? *(2 points)*

   (c) Now suppose we would like to select a model based on forecast performance. One approach is to perform time series cross-validation. (Note this model is coded up for you below).

   Suppose our objective is to predict the data for the next year. Perform the following cross-validation scheme for both models created above.

   i. For each year in $\{1985, 1986, ..., 1995\}$,
      1. Train the models based on all data before the selected year.
      2. For each of the models, generate forecasts for the 12 months in the selected year and compute the sum of squares of errors of the forecasts.

   ii. For each model, report the overall cross-validated root mean squared error (RMSE). To do this, simply divide the model's total sum-of-squares by the total number of forecasts made, then take the square root. These are the cross-validation scores of the models.

   iii. Report the cross-validation scores. Which model yields the smallest cross-validation score?

   Two hints. First, to avoid numerical issues, when fitting the linear regression model (Model 2), you might consider t ranging from 1 to the number of observations in the training set instead of starting at 1948. Second, there is a code outline on the next page that you can use for time series cross validation. Within it is code that well help with fitting the parametric signal model too.

```
start_year <- 1985
end_year <- 1994
sum_squared_errors <- c(model1=0, model2=0)
for (year in start_year:end_year) {
  train_set <- ???
  test_set <-  ???
  #
  forecast1 <- sarima.for(train_set, n.ahead=12, pdqSPDQ???)$pred
  #
  x_train = 1:length(train_set)
  model2 <- lm(train_set ~ poly(x_train, degree=3, raw=TRUE) + I(factor(x_train%%12)))
  x_test = (length(train_set) + 1):(length(train_set) + 12)
  test_matrix <- model.matrix( ~ poly(x_test, degree=3, raw=TRUE) + I(factor(x_test%%12)) )
  forecast2 <- test_matrix %*% model2$coefficients
  #
  sum_squared_errors[1] = sum_squared_errors[1] + sum((forecast1 - test_set)^2)
  sum_squared_errors[2] = ???
}
```

(5 points)

2. **Spectral Density of AR Processes**

Let W_t be a white noise process with variance 1. Consider the AR(2) process:

$$(1 - .9B^2)X_t = W_t$$

a. Compute the transfer and power transfer functions associated with the AR polynomial $(1 - .9B^2)$. Also, compute the spectral density $f_X(\lambda)$. *(3 points)*

b. Plot the spectral density $f_X(\lambda)$. Do you think X_t will oscillate? If so, what period? *(1 point)*

c. Simulate X_t for 50 time steps. Is the simulation consistent with your answer to (b)? *(1 point)*

d. Consider the linear filter with weights $a_{-1} = a_0 = a_1 = \frac{1}{3}$; $a_j = 0$ otherwise. Let Y_t be the time series obtained by applying this filter to X_t. Compute the transfer function, power transfer function, and spectral density $f_Y(\lambda)$. *(3 points)*

e. Plot the spectral density $f_Y(\lambda)$. Do you think Y_t will oscillate? If so, what period? *(1 point)*

f. Simulate Y_t by applying the filter from (d) to your simulated X_t from (c). Is the simulation consistent with your answer to (d)? *(1 point)*