MSDS 596 Regression & Time Series Fall 2018

Homework 5

Due: Tues 11/27/18 @ 6:40pm rutgers.instructure.com/courses/17597

Problem 1. Suppose we have the following model

$$r_t = 1 + 0.6r_{t-1} - 0.4r_{t-2} + a_t,$$

where $\{a_t\}$ is a normal white noise series with mean zero and variance 0.02.

(a) What is the mean of the time series r_t?

(b) Find the lag-1, lag-2 and lag-3 autocorrelations of r_t.

(c) Simulate a time series of length 2000 from this model. Create a time series plot and a sample autocorrelation plot. Report the values of the lag-1, lag-2 and lag-3 sample autocorrelations.

(d) Create a sample partial autocorrelation plot, and report the values of the lag-1, lag-2 and lag-3 sample partial autocorrelations.

Problem 2. Rewrite the model

$$(1 - .3B)(1 - .5B)(1 - .6B + .4B^2)(r_t - 1.3) = a_t$$

in its original form that does not involve the back-shift operator B. Here, $\{a_t\}$ is a white noise series. What kind of time series model does r_t follow, and what's the order of the model?

Problem 3. This problem follows Problem 3 of Homework 4. Read the dataset `hawaii-new.dat` and continue with the following analysis.

(a) Perform a log transformation of the `total` series, and fit a trend-seasonal model to the log transformed total series. Plot the fitted values with the log transformed data and plot the de-trend-de-seasoned series. Comment on the estimated coefficients of the seasonal factors.

(b) Use the trend-seasonal model in the previous question to predict the total number of tourists (in log) who will visit Hawaii each month in 1996, assuming the noises in the trend-seasonal model are i.i.d. Plot your predictions in a visually distinct way, alongside the last three years of the original data.

(c) Use functions `decompose()` and `stl()` to perform simultaneous decomposition of the original `total` series.

Problem 4. This problem follows Problem 4 of Homework 4. Read the dataset `lt.txt` and continue with the following analysis.

(a) Re-run your Kalman filter program, and calculate the exact log likelihood of the data.

(b) Assume the observations y_5, y_{100} and y_{165} are missing, calculate the exact log likelihood of the incomplete data.

(c) Pretend that you do not know σ_e^2 and σ_η^2, but you do have the prior knowledge that $s_0 \sim N(0.2, 2.25)$, use the `dlm` package to find the MLE of σ_e^2 and σ_η^2.

(d) Using the MLE obtained from above to reproduce the plots for 1) predicted state variables $s_{t|t-1}$, 2) filtered state variables $s_{t|t}$, and 3) smoothed state variables $s_{t|T}$, as you did in Homework 4 Problem 4(b) through (d).

(e) Repeat part (c), using the `R` function `StructTS()`, with `type="level"` for local trend model. Compare these variance estimates with those obtained from part (c).

Problem 5. The monthly U.S. unemployment rate dataset spans January 1948 to March 2009, and has been seasonally adjusted by the Federal Reserve Bank of St Louis. Read the dataset from `http://faculty.chicagobooth.edu/ruey.tsay/teaching/fts3/m-unrate.txt`. Monthly unemployment rate is recorded in the 4^{th} column. Take a difference of the original series and analyze this differenced series as follows.

(a) Build a autoregressive model of order 12 for the series. Report the estimated parameters, with corresponding standard errors.

(b) Calculate the estimate $\hat{\phi}_0$ of the intercept ϕ_0.

(c) Perform a test for each autoregressive coefficient, under the null hypothesis that it is equal to zero. Use the 5% level.

(d) Restrict the non-significant coefficients as zero, and refit the model. (Use the `fixed` option within the `arima()` function; do not restrict the intercept term). Report the estimated parameters, with corresponding standard errors.

(e) Perform the Ljung-Box test on the residual time series at a lag that you deem appropriate. Check for causality of the AR model.

(f) Refit the model using the data until December 2007, which is the forecast origin. Forecast the unemployment rates for the next 15 months.

(g) In the same graph, plot the true observations for the next 15 months, the forecasts, and the prediction intervals. Make sure to label the time axis by real time instead of integer index.