CSC425 – Time series analysis and forecasting
Homework 3

PROBLEMS

Problem 1 [15 pts]

Consider the following AR(2) time series process: $v_t = 0.10 + 0.4v_{t-2} + a_t$, where $\{a_t\}$ is a white noise series with mean zero and constant variance σ^2=0.02. Note that the AR(2) process has coefficient ϕ_1=0.

a) What is the mean of the time series v_t?
b) Determine if the AR(2) model is stationary by using the characteristic polynomial.
c) Assume that v_{100}= - 0.01 and v_{99} = 0.02. Compute the 1-step and 2-step ahead forecasts of the AR(2) series at the forecast origin t=100.
d) Compute the lag-1 and lag-2 autocorrelations of v_t
e) Use the code from the class in BuildingSeries.R to build a realization of this AR(2) process of 1000 time periods (use an initial value of 0 like we did). Then calculate the mean and first two autocorrelations. How closely do they match your theoretical calculations?

Problem 2 [15 pts]

Consider the following MA(2) time series process: $v_t = 5 + a_t - 0.5a_{t-1} + 0.25a_{t-2}$, where $\{a_t\}$ is a Gaussian white noise series with mean zero and constant variance σ^2=0.025.

a) What are the mean and variance of the time series v_t?
b) Is the MA(2) model is stationary? Explain.
c) Assume that a_{100}= - 0.01 and a_{99} = 0.02 and a_{98}=-0.04. Compute the 1, 2 and 3-step ahead forecasts of the MA(2) series at the forecast origin t=100.
d) Identify the number of autocorrelation coefficients for the MA(2) model that are not zero and compute them.
e) Use the code from the class in BuildingSeries.R to build a realization of this AR(2) process of 1000 time periods (use an initial value of 0 like we did). Then calculate the mean and first two autocorrelations. How closely do they match your theoretical calculations?

Problem 3 Extra Credit [10 pts]

Consider the following MA(3) time series process: $v_t = 5 + a_t - 0.5a_{t-1} + 0.25a_{t-2} - 0.1a_{t-3}$, where $\{a_t\}$ is a Gaussian white noise series with mean zero and constant variance σ^2=0.025.

a) Derive the formula for the lag-1 autocorrelation for the time series v_t like we did in class. You can use for a model, the computations for the MA(2) series from the lectures.
b) Use your formula to calculate the lag-1 autocorrelation for v_t?
c) Use the code from the class in BuildingSeries.R to build a realization of this MA(3) process of 1000 time periods (use an initial value of 0 like we did). Then calculate the mean and first two autocorrelations. How closely do they match your theoretical calculations?

Problem 4 [15 pts]

Continue your investigation into the NAPM.csv time series from the last homework.

a) Use auto.arima to compute the AIC's best fit model. Report the model coefficients with significance. Note that this series actually contains a weak seasonality, but we do not want to consider that here, so add the parameter "seasonal=F" to your auto.arima call.
b) Use auto.arima but change the fit criterion to the BIC (see the "ic" parameter in the auto.arima function. You can give it the value "BIC"). Does this change the fitted model at all? Compare the number of parameters and the sigma squared of these two models with the one you computed in homework 2.
c) Choose a model to continue your analysis and justify your choice.
d) Compute up to 5-step ahead forecasts with origin at the end of the data, i.e. 12/01/2015. Write down the forecasts and their 95% prediction intervals.
e) Plot the 10-step ahead forecasts and discuss whether the forecasts exhibit a trend that is consistent with the observed dynamic behavior of the process
f) A PMI reading above 50 percent indicates that the manufacturing economy is generally expanding; below 50 percent that it is generally declining. Do the model forecasts predict that manufacturing economy is generally expanding or contracting?
g) What do the model forecasts converge to?

Problem 5 [20 pts]

The Industrial Production Index (INDPRO) is an economic indicator that measures real output for all facilities located in the United States manufacturing, mining, and electric, and gas utilities. Since 1997, the Industrial Production Index has been determined from 312 individual series. The index is compiled on a monthly basis to bring attention to short- term changes in industrial production. Growth in the production index from month to month is an indicator of growth in the industry. Monthly percentage changes of the INDPRO index from January 1998 to December 2016 were obtained from the St Louis Federal Reserve Bank and are defined as Xt = $(p_t-p_{t-1})/p_{t-1}$ where p_t is the monthly Industrial Production Index. The dataset contains two variables: date, rate, where rate is the INDPRO index growth rate series. The following problem focuses on building a TS model for the INDPRO index growth rate series Xt.

a) Import the data and create a time object for *rate*.
b) Create the time plot of the index growth rate v_t and analyze series for stationarity, trends and strong seasonality?
c) Analyze the distribution of v_t. Can you assume that v_t is normally distributed?
d) Conduct both a Dickey-Fuller and an KPSS unit root test to determine wither the series is non-stationary and continue with the following parts using either the original series if it is stationary and the differenced series if it is not.
e) Analyze the ACF, PACF and EACF functions and discuss the following questions (a yes or no answer is not sufficient)
   - Does the TS exhibits a stationary behavior?
   - Does the process show more of an AR, MA or ARMA behavior?
f) Use your answers from the last part to choose an order and fit an ARIMA models until you get white noise residuals as tested by the Ljung-Box test (call your final model M1)

- Examine the significance of the model coefficients and analyze the residuals to check adequacy of the model.
- Write down the model expression and discuss if this is a good model for the data.

g) Apply the BIC criteria using the auto.arima(xvar, ic="bic") function, and fit the model suggested by the BIC criterion. (call this one M2)
   - Examine the significance of the model coefficients and analyze the residuals to check adequacy of the model.
   - Write down the model expression and discuss if this is a good model for the data.

h) Plot the forecasts for the models M1 and M2 and compare their behavior. Are the forecasts similar in behavior? Do you notice any major difference?

i) At the end of this analysis, which model would you recommend for use in decision making going forward? Justify your answer and discuss the chosen model and the behavior captured by it.

Problem 6 [20 points]

This problem uses the dataset *consump.csv* that contains 3 U.S. macroeconomic variables for the period of 2000 to 2013. The goal of this problem is to analyze and predict the variable "pers_inc" representing monthly values of real personal income.

.

a) Create a time series for the *pers_inc* variable and make a time plot of series and describe its behavior over time. Does it appear to be non-stationary, have a trend or very strong seasonality? Is it multiplicative and does it need to be log-transformed?

b) Plot the ACF, PACF and EACF function (20 lags) of *pers_inc* and discuss whether the time series shows a non-stationary behavior

c) Plot the ACF, PACF and EACF function (20 lags) of the first difference of *pers_inc* and discuss whether the differenced time series shows non-stationary behavior. Is there any evidence of over-differencing (see the discussion at the end of week 4's synchronous session) in the ACF plot of the first difference?

d) Use Dickey Fuller to test the hypothesis of unit-root non stationarity. Which version of the test is appropriate given your observations in a)?

e) Use the results of d) and the information from the plots in b) and c) to choose an order for an ARIMA(p, d, q) fit, and also decide whether a drift is appropriate. Create the fit and report the coefficients along with their significance.

f) Fit the selected ARIMA model, and use residual analysis to check if the model captures the dynamic behavior of monthly real personal income.

g) Plot a 20 step ahead forecast for the end of the series and evaluate its performance.

h) Use auto.arima to identify the order of the "best" ARIMA(p,d,q) model. Does the selected model include a drift? How does it compare with your model. If the model differs from the one you chose, compare the two models for their residual autocorrelation, including a Ljung-Box test on each.

"Reflection" Problem [5 pts]

Post a message on the discussion board reflecting on the topics in weeks 3 and 4. Indicate the assignment in this module you found to be the easiest, the one you found to be the hardest, and why. I created a new Thread called “Lectures 3 & 4 and Homework 3”.

Submission instructions

Submit the homework at the Course Web page http:d2l.depaul.edu in the Dropbox page. Keep a copy of all your submissions!

1. Submit your answers in a single pdf document. Make sure to explain in detail your analyses and include relevant output and graphs. Paste the relevant portions of your code in your document along with your plots in a **well organized manner**. Your presentation should be clean and your answers to each part should be **clearly visible.**
2. If you have questions about the homework, email me BEFORE the deadline.
3. Late assignments will be penalized 25% per the syllabus.
4. Assignments submitted five days after the deadline will not be accepted.