PROJECT WORK IN SF2943: TIME SERIES ANALYSIS (2017)

PIERRE NYQUIST

This course has mandatory project work as an essential learning activity. The course aims to provide you with sufficient skills to analyze time series data through a combination of theoretical and practical work. The project work addresses the latter. The problem statements are somewhat vague by design - you are required to use good judgement in deciding on what techniques and methods to use in solving them. You are also encouraged to let your curiosity guide you in your investigations and possibly explore other aspects of the problems than those explicitly stated here.

1. General instructions

The project work is done in groups of 1-4 students and is evaluated on the basis of your written report. Email the names and email addresses of your group member (one email per group) to pierren@kth.se no later than April 4.

You may use whatever software you like (e.g., Matlab, R, ...), use available functions for time series analysis, or write your own functions. Whatever you choose, you are required to state precisely what you are using and what the specific function does; it is not enough to only provide the name of a function and its output. In Problem 6 you are required to use Quantlab (see separate documentation). For the report there will be a LaTex template available on the course website.

There are two mandatory project discussion seminars (dates in Section 2). During the seminars you will discuss the project work with other groups and active participation is required for obtaining a “Pass” on the project-part of the course. For the first discussion seminar, in order to participate you are required to bring plots and preliminary results that will make the discussion meaningful for all parties. For the second seminar, you are required to bring a preliminary version of your project report, with at least partial solutions to problems 1- 6 in Section 3.

In addition to the two discussion seminars there are two mandatory presentation seminars. For the first seminar you must be prepared to present and discuss your solutions to Problems 1- 6. You will be informed one day in advance what part of the project you will be asked to present. If one member of the group gives a presentation that indicates that he/she does not fully understand the work of the group, that will affect the evaluation of the entire group.

The second presentation seminar concerns Problem 7: You are asked to study at least three old exams and select two problems that you find particularly interesting and relevant for the course. For the presentation seminar you should be prepared to present solutions and motivate your choice of problems. Two days before the seminar you are required to hand in a *brief* (1/2-1 page) motivation for your selection, what the problems test and if possible suggest modifications that make the problems more relevant or interesting.

2. Dates and deadlines

- April 7: A list of names and email addresses for your group members to pierren@kth.se.
- April 18: First discussion seminar.
- May 2: Second discussion seminar.
- May 5 **12:00**: Hand in written report.
- May 10: First presentation seminar.
- May 15: Hand in report on previous exam problems.
- May 17: Second presentation seminar.

3. Problems

Problem 1: White noise. When considering a time series data set, how would you decided whether or not it is plausible that it is a realization of white noise, or even a sequence of independent and identically distributed (iid) random variables?

Simulate iid samples of varying size n and compute sample autocorrelations $\hat{\rho}(h)$for different lags h. Investigate the claim that, for a sufficiently large sample size n, the $\hat{\rho}(h)$ for different h are approximately independent and have a Gaussian distribution with mean zero, variance $1/n$. Perform relevant simulations of iid samples, statistical tests and analyze the results.

In addition to the claim regarding the sample autocorrelations, apply some other methods discussed in Section 1.6 of the textbook on iid samples and study their performance.

Repeat the analysis with the iid sample replaced by a low-order AR and MA process with small coefficients. How large must the coefficients be for the different methods to detect deviations from white noise? What is the effect of the sample size on the methods? What can you say about the possibility to detect a small trend in the data?

Problem 2: Parameter estimation. Consider the AR(2) time series model

$$X_t - 1.3X_{t-1} + 0.65X_{t-2} = Z_t, \quad \{Z_t\} \sim WN(0, 280), \tag{1}$$

where the white noise sequence consists of independent Gaussian random variables (hence, it is really an $IID(0, 280)$ sequence). The task is to evaluate methods for parameter estimation in such a model.

Simulate m samples of size n from the $AR(2)$ model (1), with independent Gaussian random variables Z_t. For each of the m samples, use Maximum Likelihood and Yule-Walker estimation for estimating the model parameters. Present the estimation methods in detail. For each method, present a scatter plot of the m pairs of estimated coefficients. Which methods appears to be most accurate? Which method produces the smallest one-step mean-squared prediction error? Compare the histograms - remember to use the same x-axis - of the one-step prediction errors. Lastly, investigate whether the residuals from the fitted model are uncorrelated and have the same distribution as the original noise sequence $\{Z_t\}$.

Problem 3: Prediction.
a) Simulate m samples of size n from the $AR(2)$ model (1) with independent Gaussian Z_ts. For each sample fit both an $AR(2)$ and an $AR(10)$ model. Compare how well the two fitted models do one-step and three-step predictions.

b) Consider the $AR(1)$ process

$$X_t = 0.8X_{t-1} + Z_t, \quad \{Z_t\} \sim WN(0,1). \tag{2}$$

Simulate m samples of size n of the process $\{X_t\}$, with independent and Gaussian Z_ts, and for each sample fit an $AR(1)$ and an $MA(10)$ model to the data. Compare how well the two fitted models do one-step and three-step predictions. Motivate why we may attempt to use an MA model for this type of process X_t.

c) Consider the $AR(1)$ process (2) with independent Z_ts and with $c+Z_t$ having a log-normal distribution for an appropriate choice of c. Simulate m samples of size n from the model (2). For each sample fit an $AR(1)$ model. Make a scatter plot of the estimated parameter pair corresponding to the coefficient in the AR model and the standard deviation of the white noise sequence. How well does the fitted model do one-step predictions? Compare to the findings in (b).

Problem 4: Model selection. Consider the time series data sets available on the course website. Your task is to find an appropriate model for the time series data.

Whenever necessary, transform the data into a time series that appears to be a realization of a stationary time series model. How do you know whether you have succeeded in removing possible trends and seasonal components? Choose a stationary time series model for the (possibly transformed) time series data; present your analysis for deciding on the appropriate model class out of $AR(p)$, $MA(q)$ and $ARMA(p,q)$. What are appropriate values for p and/or q? Estimate the parameters of the chosen time series model. Elaborate on difficulties and alternative approaches.

Problem 5. Choose time series data that you would like to analyze. Present the time series and explain your interest in the data. Analyze the time series using appropriate techniques from your solution to Problem 4, and any other techniques that seem relevant for the specific time series. Present your analysis and describe your findings.

Problem 6: GARCH models. See separate problem statement (available on the course website).

Problem 7: Old exam problems. Go over three previous exams in the course and choose two problems that you find particularly relevant. Present solutions to the problems, explain what the problems test and motivate their relevance as exam problems. If possible, suggest modifications that would improve the problems.