Statistics 153 (Introduction to Time Series) Homework 1

Due on **Thursday, September 13th**, 2018

August 31, 2018

General comments: Each homework is worth 20 points. You do not receive individual letter grades for your homework. Your final grade will be a weighted average of your homework (30%), Midterm 1 (15%), Midterm 2 (15%), and Final (40%). There will be 6 homework assignments in total, roughly posted every second week. You will be allowed to drop one of the homework assignments for your final assessment. Usually, homework gets **posted on (or before) a Tuesday** and will be **due at the beginning of the Tuesday lecture** the week after. For this first homework, we make an exception: It is due on the beginning of the **Thursday** lecture the week after. Late assignments that are handed in after 3:50pm will not be accepted! Assignments should include your full name and the GSI's name of the lab section you are enrolled in. Your homework must be stapled when you hand it in. You are encouraged to work in small groups on homework problems. However, you must write up the solutions on your own, and you must never read or copy the solutions of other students. Similarly, you may use books or online resources to help solve homework problems, but you must credit all such sources in your writeup and you must never copy material verbatim. Any student found to be cheating risks automatically failing the class and being referred to the Office of Student Conduct. In particular, copying solutions, in whole or in part, from other students in the class or any other source without acknowledgment constitutes cheating.

Computer exercises:

1. Download the google trends time series dataset for the query *microsoft*.

   (a) Estimate the trend function in the data by fitting a parametric curve to the data. Provide a plot of the original data along with the corresponding trend estimate. Also provide a time plot and correlogram of the residuals. Comment on each of these plots.

   (2 Points)

   (b) Estimate the trend by smoothing. Explain reasons behind your choice of the smoothing parameter. Once again, provide a plot of the original data along with the corresponding trend estimate. Also provide a time plot and correlogram of the residuals. Comment on each of these plots.

   (2 Points)

(c) Difference the data and provide a plot of the differenced data. Is there any trend in the differenced data? Also plot the correlogram of the differenced data and comment on it.

(2 Points)

(d) Estimate the trend function in the data using isotonic estimation. Provide a plot of the original data along with the corresponding trend estimate. Also provide a time plot and correlogram of the residuals. Comment on each of the plots.

(2 Points)

Theoretical exercises:

2. While analyzing their annual sales data (number of car sales per year) for the past 30 years, a car company found that after taking three successive differences, the resulting data had a mean of 2051 and looked like white noise. If the actual data for the past three years were 2017 - 61214, 2016 - 52574, and 2015 - 39381. What would be a reasonable forecast for sales in 2018? Explain.

(1 Points)

3. For white noise $X_1, \ldots, X_n$ with n sufficiently large, what is (approximately) the probability that if you plot in R the correlogram for the first 100 lags $r_1, \ldots, r_{100}$ **at least** 3 **of the** r_k**'s** lie outside of the blue 5% confidence band (i.e., for at least three r_k's it holds that $|r_k| \geq 1.96/\sqrt{n}$)?

(2 Points)

4. Consider the stochastic trend model

$$X_t = m_t + Z_t, \quad t = 1, \ldots, n$$

where Z_t is a white noise process (with variance σ_Z^2) and m_t is a stochastic trend which follows a random walk with drift model

$$m_t = \delta + m_{t-1} + W_t,$$

where W_t is another white noise process (with variance σ_W^2), independent of Z_t. Let $m_0 = 0$.

(a) Find the mean function $\mu(t) = \mathbb{E}(X_t)$, autocovariance function $\gamma(s,t) = \mathrm{Cov}(X_t, X_s)$, and autocorrelation function $\rho(s,t) = \gamma(s,t)/\sqrt{\gamma(t,t)\gamma(s,s)}$ of $\{X_t\}$. To this end, first show that the model can be written as $X_t = \delta t + \sum_{k=1}^{t} W_k + Z_t$.

(3 Points)

(b) A time series process is denoted as *weakly stationary* if the mean function $\mu(t)$ is constant (i.e., does not depend on t) and the autocovariance function $\gamma(s,t)$ depends on s and t only through their difference $|s-t|$. Is the process X_t defined above weakly stationary? Explain.

(1 Points)

(c) Show that $\rho(t-1,t) \to 1$ as $t \to \infty$. What is the implication of this result?

(1 Points)

(d) Suggest a transformation to make the series X_t weakly stationary and prove that the transformed series is weakly stationary.

(2 Points)

5. Consider a monthly time series dataset for which we believe that the model $X_t = (at+b)s_t + Z_t$ is appropriate where s_t is a seasonal function of known period d i.e., $s_{t+d} = s_t$ and Z_t is white noise. What would be a way to difference the data in order to eliminate both trend and seasonality (such that the resulting differenced process has a mean function which does not depend on t) and why does it work?

(2 Points)