

# STAT 153 – Introduction to Time Series

## Homework 1

Due Wednesday **September 8**, 2021, 3 pm

**Computer exercises.** The dataset for these exercises is on bCourses.

1. Load the `iPod.csv` file which contains the Google trends data concerning the query *iPod* from 2004 through most of 2019.
  - (a) (10 points) Estimate the trend function in the data by fitting a *linear* parametric trend line to the data. Provide a plot of the original data along with the corresponding trend estimate. Also provide a time series plot of the residuals. Comment on each of these plots.
  - (b) (10 points) Estimate the trend function in the data by fitting a *more complex* parametric curve to the data (essentially everything is more complex than linear). Provide a plot of the original data along with the corresponding trend estimate. Also provide a time series plot of the residuals. Comment on each of these plots.
  - (c) (10 points) Which of these two models' forecasts would you trust the most? The function may fit well in-sample ( $t=1, \dots, 189$ ), but it may not forecast well ( $t=190, 191, \dots$ ). Plot the two trend models' forecasts out to the end of 2020 (up to  $t=204$ ). Comment on which of these provides a reasonable forecast, or if both are poor.

**Theoretical exercises.**

2. Let  $X_1, \dots, X_n$  be white noise with  $n$  sufficiently large. Recall that the ACF correlogram in R has blue 95% confidence bands at  $\pm 1.96/\sqrt{n}$ .
  - (a) (10 points) If we plot  $H$  sample autocorrelations  $(r_k, k \geq 1)$ , how many of them do we expect to exceed (stick out of) the confidence bands?
  - (b) (10 points) What is the probability that a single sample autocorrelation  $r_k, k \geq 1$ , significantly exceeds the confidence bands, to the amount of 20% the distance from 0 to the confidence band? Or in notation:  $P(|r_k| \geq [1.2 \times 1.96/\sqrt{n}]) = ?$
  - (c) (10 points) What is (approximately) probability that for the correlogram of the first 100 lags  $r_1, \dots, r_{100}$  **at least 4 of the**  $r_k$ 's lie outside of the blue confidence bands (i.e., for at least four  $r_k$ 's it holds that  $|r_k| \geq 1.96/\sqrt{n}$ )?

3. (a) (15 points) Let  $\{X_t\}$  and  $\{Y_t\}$  be weakly stationary sequences. Also assume that  $X_r$  and  $Y_s$  are uncorrelated for all  $r$  and  $s$ . Show that  $\{X_t + Y_t\}$  is weakly stationary with autocovariance function equal to the sum of the autocovariance functions of  $\{X_t\}$  and  $\{Y_t\}$ .

- (b) Let  $\{Z_t\}$  be iid  $\mathcal{N}(0, 1)$  noise and define

$$X_t = \begin{cases} Z_t & \text{if } t \text{ is even} \\ (Z_{t-1}^2 - 1)/\sqrt{2} & \text{if } t \text{ is odd} \end{cases}$$

- i. (10 points) Is  $\{X_t\}$  white noise? Why or why not?
- ii. (5 points) Is  $\{X_t\}$  IID noise? Why or why not?
- iii. (5 points) Is  $\{X_t\}$  weakly stationary? Why or why not?
- iv. (5 points) Is  $\{X_t\}$  strongly stationary? Why or why not?