STAT 153 - Introduction to Time Series
Homework Two

Fall 2022, UC Berkeley

Due by 11:59 pm on 27 September 2022

Total Points = 64

1. In Figure 1, you will find two different time series datasets (Dataset One and Dataset Two) of the same length $n = 1000$. In Figure 2, you will find two periodograms (Periodogram A and Periodogram B). One of these periodograms corresponds to Dataset One and the other to Dataset Two. Identify the correct periodograms giving reasons for your answer. (**5 points**)

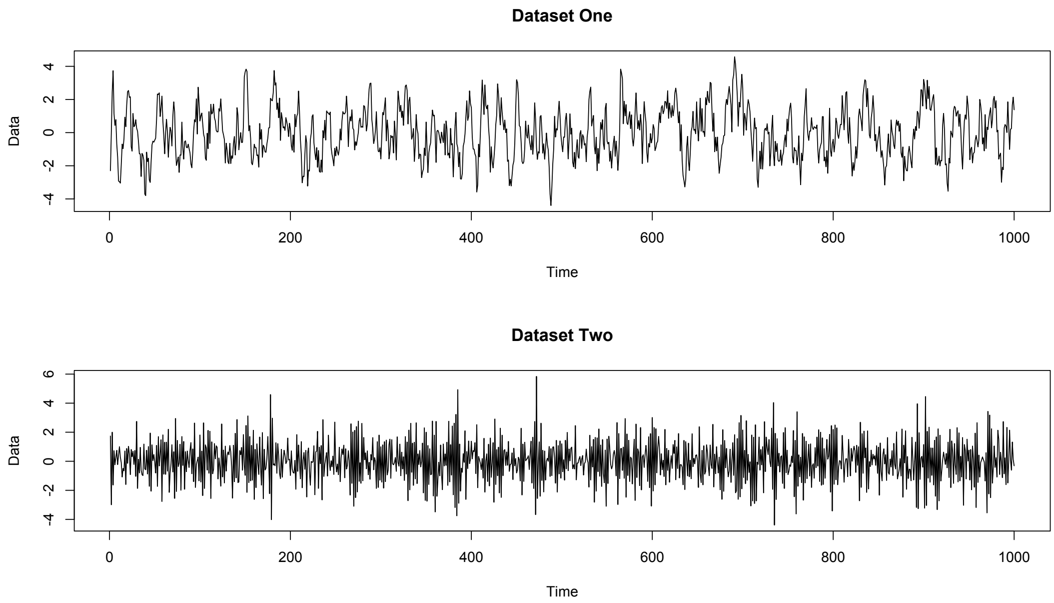

Figure 1: Two Time Series Datasets

2. Consider the dataset `lynx` that is available in base R. This gives the annual numbers of lynx trappings for 1821-1934 in Canada. Type `help(lynx)` to learn more about the dataset.

   a) Plot the periodogram of the data and comment on its notable features. (**3 points**).

   b) To this dataset, fit the model:

$$Y_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t) + Z_t \qquad \text{with } Z_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2) \tag{1}$$

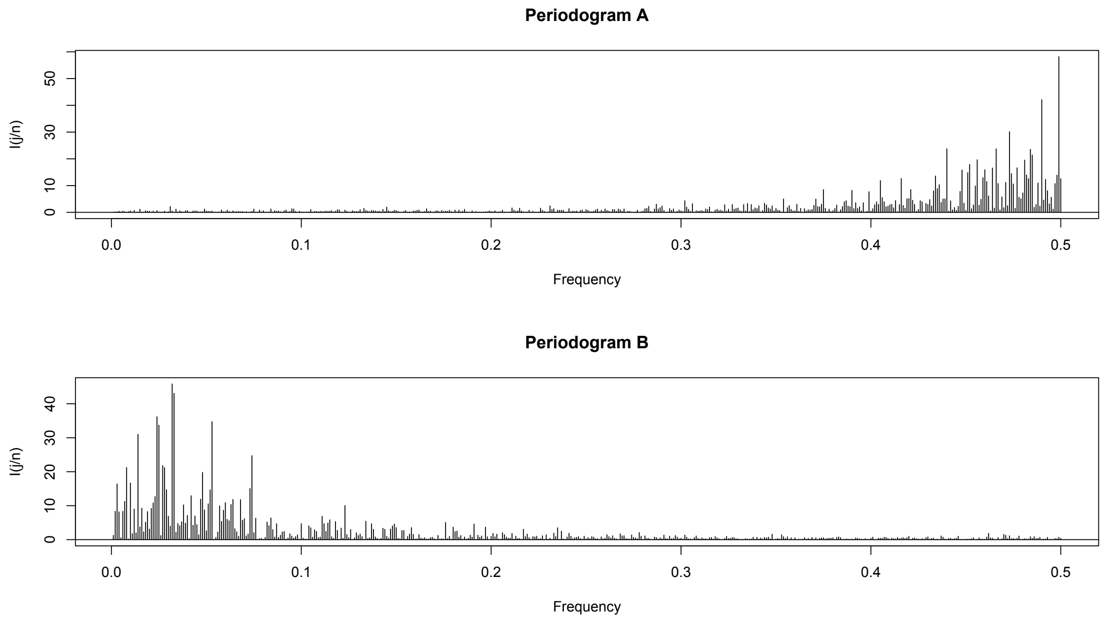

Figure 2: Two Periodograms

Provide a point estimate and 95% uncertainty interval for the frequency f and the corresponding period of oscillation. (**4 points**).

c) Provide point estimates and 95% marginal uncertainty intervals for $\beta_0, \beta_1, \beta_2$ and σ (**6 points**).

d) Comment on whether (1) is a good model for this dataset. (**2 points**)

3. Download the Google Trends Data (for the United States) for the query *mask*. This should be a monthly time series dataset that indicates the search popularity of this query from January 2004 to September 2022. To this data, fit the single change point model:

$$Y_t = \beta_0 + \beta_1 I\{t > c\} + Z_t \qquad \text{with } Z_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2). \tag{2}$$

a) Provide a point estimate and 95% uncertainty interval for the changepoint parameter c. Explain whether your answers make intuitive sense in the context of this dataset (**4 points**).

b) Provide point estimates and 95% marginal uncertainty intervals for the pre-changepoint mean level $\mu_0 := \beta_0$ and the post-changepoint mean level $\mu_1 := \beta_0 + \beta_1$. (**4 points**).

c) Comment on whether (2) is a good model for this dataset. (**2 points**)

4. Download the Google Trends Data (for the United States) for the query *golf*. This should be a monthly time series dataset that indicates the search popularity of this query from January 2004 to September 2022. To this data, fit the model:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 \cos(2\pi f t) + \beta_4 \sin(2\pi f t) + Z_t \qquad \text{with } Z_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2). \tag{3}$$

a) Provide a point estimate and a 95% uncertainty interval for the unknown frequency parameter f. (**4 points**)

b) On a scatter plot of the data, plot your best estimate of the fitted function:

$$t \mapsto \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 \cos(2\pi f t) + \beta_4 \sin(2\pi f t)$$

along with appropriate uncertainty quantification. (**5 points**).

c) Comment on whether model (3) is appropriate for this dataset. (**2 points**).

5. Download the FRED dataset on Total Construction Spending in the United States from `https://fred.stlouisfed.org/series/TTLCONS`. This gives monthly seasonally adjusted data on total construction spending in the United States in millions of dollars from January 1993 to July 2022. To this dataset, fit the model:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 (t - s_1)_+ + \beta_3 (t - s_2)_+ + Z_t \tag{4}$$

with $Z_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. The unknown parameters in this model are $\beta_0, \beta_1, \beta_2, \beta_3, s_1, s_2, \sigma$.

a) Provide point estimates and 95% uncertainty intervals for the change of slope parameters s_1 and s_2. (**5 points**)

b) On a scatter plot of the data, plot your best estimate of the fitted function:

$$t \mapsto \beta_0 + \beta_1 t + \beta_2 (t - s_1)_+ + \beta_3 (t - s_2)_+$$

along with appropriate uncertainty quantification. (**5 points**).

c) Comment on whether model (4) is appropriate for this dataset. (**2 points**).

6. Download the google trends time series dataset for the query *playoffs* (download the trends for the United States and not worldwide). This should be a monthly time series dataset that indicates the search popularity of this query from January 2004 to September 2022.

a) Plot the data and observe that the scale of variability increases with time. To fix this, take the logarithm of the data. Plot the logarithm and comment on whether the variability can now be assumed to be constant over time. (**2 points**)

b) The logarithmed data have an increasing trend and a periodic component superimposed on the trend. To get an idea of the frequencies present in the periodic component, fit a linear trend model to the logarithmed data, compute the residuals and then plot the periodogram of the residuals. Comment on the location and size of the main spikes in the periodogram. (**3 points**)

c) To the logarithmed data, fit the model

$$y_i = g(t_i) + Z_i \qquad \text{where } Z_i \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$$

and

$$\begin{aligned} g(t) = \beta_0 &+ \beta_1 t \\ &+ \beta_2 \cos(2\pi f_1 t) + \beta_3 \sin(2\pi f_1 t) \\ &+ \beta_4 \cos(2\pi f_2 t) + \beta_5 \sin(2\pi f_2 t) \\ &+ \beta_6 \cos(2\pi f_3 t) + \beta_7 \sin(2\pi f_3 t). \end{aligned}$$

Treat $\beta_0, \beta_1, \ldots, \beta_7, f_1, f_2, f_3, \sigma$ as unknown parameters and estimate them from data. Plot your estimate of g (and appropriate uncertainty indicators) along with the actual data. (**6 points**)