STAT 153 & 248 - Time Series
Homework Four

Spring 2025, UC Berkeley

Due by 11:59 pm on 14 April 2025
Total Points = 85 (STAT 153) and 103 (STAT 248)

1. Consider the EEG dataset from `https://physionet.org/content/eegmmidb/1.0.0/`, which we previously used in Lecture 15. Access the data for the first subject (located in the `S001` folder). Download the following three files:

   - `S001R01.edf`
   - `S001R02.edf`
   - `S001R05.edf`

   These files contain EEG recordings from 64 channels (electrodes) under different conditions:

   - `S001R01.edf`: Eyes open
   - `S001R02.edf`: Eyes closed
   - `S001R05.edf`: Performing a specific task (see the website for task details)

   Focus your analysis on the data from **channel 60**, which will yield three time series—one for each condition.

   a) Plot the three time series (eyes open, eyes closed, and task). Comment on any noticeable differences between them based on the visualizations. (**2 points**)

   b) For each of the three time series, estimate the power spectral density (PSD). Plot both the periodogram and the estimated PSD on a logarithmic scale (separately for each time series). Clearly describe the method used for estimating the spectral density including the regularization technique and choice of the tuning parameter. (**6 points**)

   c) Plot all three PSDs on the same log-scale plot, using different colors for each condition. Discuss the differences in spectral properties among the three conditions (eyes open, eyes closed, and task). (**4 points**)

2. Download the FRED dataset on US population from `https://fred.stlouisfed.org/series/POPTHM`. This is a monthly dataset (units are thousands) and is not seasonally adjusted.

   a) Fit an AR(3) model to this dataset. Report estimates of the ϕ-parameters along with their standard errors. (**3 points**)

b) Use the fitted model to predict the data for 60 months immediately succeeding the last month in the dataset. Plot these predictions and uncertainty indicators along with the original data. Do these predictions make intuitive sense? (**4 points**)

c) Suppose we want to predict the US population for the months **preceding** January 1959. For this purpose, fit the model:

$$y_t = \alpha_0 + \alpha_1 y_{t+1} + \alpha_2 y_{t+2} + \alpha_3 y_{t+3} + \epsilon_t$$

for $t = 1, 2, \ldots, n-3$ (where n is the length of the dataset). Report point estimates with corresponding standard errors for the parameters $\alpha_0, \alpha_1, \alpha_2, \alpha_3$. Compare your fitted model with the AR(3) model fitted in part (a). Are there any similarities between the two models? (**4 points**)

d) Use the model from part (c) to predict the US population for the 60 months immediately preceding January 1959. Plot these predictions along with the original data. Do these predictions make intuitive sense? (**4 points**).

3. Download the FRED dataset on "Retail Sales: Beer, Wine, and Liquor Stores" from `https://fred.stlouisfed.org/series/MRTSSM4453USN`. This is a monthly dataset (the units are millions of dollars) and is not seasonally adjusted. Separate the last 36 observations from this dataset and keep them as a test dataset. Fit the $AR(p)$ model for each $p = 1, 2, \ldots, 24$ to the training dataset and use it to predict the observations in the test dataset. Evaluate the 24 models based on the mean-squared accuracy of prediction and report the model with the best prediction accuracy. (**10 points**).

4. Download the FRED dataset on "Personal Consumption Expenditures: Services" from `https://fred.stlouisfed.org/series/PCESV`. This is a quarterly dataset. Reserve the last 24 observations (corresponding to 6 years) for testing purposes. We shall fit models on the rest of the data, and then evaluate prediction accuracy on the 24 test observations. Denote the observed data by PCE_t.

   a) Fit an AR(p) model on the training dataset PCE_t. Explain how you chose the value of p. Report estimates and standard errors of the ϕ-coefficients. (**3 points**)

   b) Use the fitted model from part (a) to predict the 24 test observations. Compare the predictions with the actual test values. Plot the original data along with the predictions and the actual test values. Comment on the accuracy of prediction. (**3 points**)

   c) Now work with $\log(\text{PCE}_t)$. Fit the AR(4) model to the training $\log(\text{PCE}_t)$. Report estimates and standard errors of the ϕ-coefficients. (**3 points**)

   d) Use the fitted model from part (c) to predict the 24 test observations. Compare the predictions with the actual test values. Plot the original data along with the predictions and the actual test values. Comment on the accuracy of prediction. (**3 points**)

   e) Now work with the differenced log data:

   $$y_t = \log(\text{PCE}_t) - \log(\text{PCE}_{t-1}).$$

   Fit the $AR(3)$ model to the training y_t. Write the fitted model for y_t as an AR(4) model for $\log(\text{PCE}_t)$. Compare this AR(4) model with the AR(4) model obtained in part (c). Are the estimated coefficients similar? (**4 points**).

   f) Use the fitted model from part (e) to predict the 24 test observations. Compare the predictions with the actual test values. Plot the original data along with the predictions and the actual test values. Comment on the accuracy of prediction. (**3 points**)

g) Now work with **double**-differenced log data:

$$\tilde{y}_t = y_t - y_{t-1} = \log(\text{PCE}_t) - 2\log(\text{PCE}_{t-1}) + \log(\text{PCE}_{t-2})$$

Fit the $AR(2)$ model to the training y_t. Write the fitted model for $\tilde{y}_t$ as an AR(4) model for $\log(\text{PCE}_t)$. Compare this AR(4) model with the AR(4) models obtained in parts (c) and (e). Are the estimated coefficients similar? (**4 points**).

h) Use the fitted model from part (g) to predict the 24 test observations. Compare the predictions with the actual test values. Plot the original data along with the predictions and the actual test values. Comment on the accuracy of prediction. (**4 points**)

i) Which model gave the best predictions for the 24 test observations? (**2 points**)

Even though models were fitted to various variables $\text{PCE}_t, \log(\text{PCE}_t), y_t, \tilde{y}_t$ in this problem, the predictions should all be converted to the original data PCE_t.

5. a) Consider the difference equation:

$$u_k - u_{k-1} + (0.5)u_{k-2} = 0 \qquad \text{for } k = 0, 1, 2, \ldots$$

for two fixed values u_0 and u_1. Show that the solution is given by: (**4 points**)

$$u_k = 2^{-k/2}\left(u_0 \cos\frac{\pi k}{4} + (2u_1 - u_0)\sin\frac{\pi k}{4}\right) \qquad \text{for } k = 0, 1, 2, \ldots.$$

b) Suppose we use the AR(2) model:

$$y_t = 3 + y_{t-1} - (0.5)y_{t-2} + \epsilon_t \tag{1}$$

to predict the future values of a time series dataset $y_1, \ldots, y_n$. Show that the predictions are given by:

$$\hat{y}_{n+i} = 6 + 2^{-(i+1)/2}\left\{(y_{n-1}-6)\cos\frac{\pi(i+1)}{4} + (2y_n - y_{n-1}-6)\sin\frac{\pi(i+1)}{4}\right\}$$

for $i \geq 1$. (**4 points**)

c) Consider the model:

$$y_t = 6 + \sum_{j=0}^{\infty} 2^{-j/2}\left(\cos\frac{\pi j}{4} + \sin\frac{\pi j}{4}\right)\epsilon_{t-j}$$

where $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. Show that y_t satisfies the AR(2) equation (1). (**4 points**)

6. Suppose $U_1, U_2, V_1, V_2 \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. Also let f_1 and f_2 be two distinct fixed real numbers. For $t = \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$, let

$$y_t = U_1\cos(2\pi f_1 t) + V_1\sin(2\pi f_1 t) + U_2\cos(2\pi f_2 t) + V_2\sin(2\pi f_2 t)$$

a) Calculate $\mathbb{E}y_t$. (**1 point**)

b) Calculate $\text{var}(y_t)$. (**2 points**)

c) Calculate $\text{cov}(y_{t_1}, y_{t_2})$ for $t_1 \neq t_2$. (**3 points**)

d) Is y_t stationary? Why or why not? (**1 point**).

7. [**This question is only for students taking STAT 248**] Consider the Personal Consumption Expenditure time series from FRED (`https://fred.stlouisfed.org/series/PCEC`) that was used in Lecture 19. Consider fitting the AR(2) model to the logarithm of the data: $y_t = \log \text{PCEC}_t$. We will reserve the last 12 observations for testing and fit the model on the remaining observations. In Lecture 19, we obtained predictions and standard errors for the fitted model for the test observations using the `get_prediction` function from statsmodels. In this problem, we shall explore a full Bayesian approach.

   a) Fit the AR(2) model to the training dataset using the OLS method (which reports t-scores). Report point estimates of the ϕ-coefficients along with standard errors. (**3 points**).

   b) Generate $N = 5000$ posterior samples $\theta^{(i)} = (\phi_0^{(i)}, \phi_1^{(i)}, \phi_2^{(i)}, \sigma^{(i)}), 1 \leq i \leq N$, for the parameters $\theta = (\phi_0, \phi_1, \phi_2, \sigma)$. (**5 points**)

   c) For each $i = 1, \dots, N$, generate $y_{n+1}^{(i)}, \dots, y_{n+12}^{(i)}$ from the conditional distribution of $(y_{n+1}, \dots, y_{n+12})$ conditioned on the observed data and $\theta = \theta^{(i)}$. (**5 points**)

   d) Use the generated posterior samples $y_{n+1}^{(i)}, \dots, y_{n+12}^{(i)}$ for $i = 1, \dots, N$ to obtain point predictions (mean of the posterior samples) and standard errors (standard deviation of the posterior samples). Compare these to the results of the `get_prediction` approach. (**5 points**).