STAT 153 & 248 - Time Series
Midterm Practice Questions

Spring 2025, UC Berkeley

Aditya Guntuboyina

March 9, 2025

1. To an observed time series dataset $y_0, \dots, y_{n-1}$, I would like to fit the model

$$y_t = \beta_0 + \beta_1 \cos\left(\frac{2\pi t}{n}\right) + \beta_2 \sin\left(\frac{2\pi t}{n}\right) + \epsilon_t \qquad \text{where } \epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2) \tag{1}$$

and $t = 0, 1, \dots, n-1$.

a) Write down Maximum Likelihood Estimates of $\beta_0, \beta_1, \beta_2$ and σ.

b) Relate the Maximum Likelihood Estimates of β_1 and β_2 to the DFT of $y_1, \dots, y_n$.

c) Now consider the alternative model

$$y_t = \beta_0 + \beta_1 \cos\left(\frac{2\pi t}{n}\right) + \beta_2 \sin\left(\frac{2\pi t}{n}\right) + \beta_3 \cos\left(\frac{4\pi t}{n}\right) + \beta_4 \sin\left(\frac{4\pi t}{n}\right) + \epsilon_t, \tag{2}$$

again with $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. Consider the following claim: For every $y_1, \dots, y_n$, the Maximum Likelihood Estimates of β_0, β_1 and β_2 are the same for both models (1) and (2). Is this claim true or false? Provide reasons for your answer.

d) Consider the following claim: For every $y_1, \dots, y_n$, the Maximum Likelihood Estimate of σ is the same for both models (1) and (2). Is this claim true or false? Provide reasons for your answer.

Solution:

a) Model (1) is a linear regression model $y = X\beta + \epsilon$ where

$$X = \begin{pmatrix} 1 & \cos(2\pi(0/n)) & \sin(2\pi(0/n)) \\ 1 & \cos(2\pi(1/n)) & \sin(2\pi(1/n)) \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & \cos(2\pi((n-1)/n)) & \sin(2\pi((n-1)/n)) \end{pmatrix} \quad \text{and } \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

The MLE of β is therefore given by $(X^T X)^{-1} X^T y$. Because of the orthogonality properties of $\cos(2\pi t/n)$ and $\sin(2\pi t/n)$ derived in class, we get

$$(X^T X)^{-1} = \begin{pmatrix} n & 0 & 0 \\ 0 & (n/2) & 0 \\ 0 & 0 & (n/2) \end{pmatrix}$$

leading to

$$\hat{\beta}_0 = \bar{y}, \quad \hat{\beta}_1 = \frac{2}{n}\sum_t y_t \cos(2\pi t/n), \quad \hat{\beta}_2 = \frac{2}{n}\sum_t y_t \sin(2\pi t/n)$$

The MLE of σ is $\sqrt{RSS/n}$ which is

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{t=0}^{n-1}\Big(y_t - \hat{\beta}_0 - \hat{\beta}_1\cos(2\pi t/n) - \hat{\beta}_2\sin(2\pi t/n)\Big)^2}$$

b) The DFT term b_1 has real part $\sum_t y_t \cos(2\pi t/n)$ and imaginary part $-\sum_t y_t \sin(2\pi t/n)$. Therefore

$$\hat{\beta}_1 = \frac{2}{n}\mathrm{Re}(b_1) \quad \text{and} \quad \hat{\beta}_2 = -\frac{2}{n}\mathrm{Im}(b_1). \tag{3}$$

c) This claim is true. The main reason is the orthogonality of sines and cosines at Fourier frequencies. Model (2) is also a linear model with a different X. Due to the orthogonality of the sines and cosines, we now have

$$(X^T X)^{-1} = \begin{pmatrix} n & 0 & 0 & 0 & 0 \\ 0 & n/2 & 0 & 0 & 0 \\ 0 & 0 & n/2 & 0 & 0 \\ 0 & 0 & 0 & n/2 & 0 \\ 0 & 0 & 0 & 0 & n/2 \end{pmatrix}$$

from which it is easy to see (using the formula $(X^T X)^{-1} X^T y$) that the MLEs for $\beta_0, \beta_1, \beta_2$ coincide in both models.

d) This claim is false. The MLE $\hat{\sigma}^2$ is $\frac{1}{n}$ times the sum of squared residuals. In Model (2), we can fit at least as well as in Model (1), because we have additional parameters β_3, β_4. Orthogonality means these new terms do not change $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$, but they can capture additional variation in $\{y_t\}$ if the data has a component at frequency $4\pi/n$. Therefore, in general, the sum of squared residuals under Model (2) will be less than or equal to the sum of squared residuals under Model (1), which in turn yields a (weakly) smaller estimate of $\hat{\sigma}^2$. Thus, for many datasets (in which the sinusoids at the second Fourier frequency explain additional variance), $\hat{\sigma}^2$ is strictly smaller in Model (2). Consequently, $\hat{\sigma}$ differs in the two models.

2. The `UKgas` dataset in R gives Quarterly Observations on the UK gas consumption from the first quarter of 1960 to the last quarter of 1986 (there are 108 observations in total). A plot of the data is given in Figure 1. Consider the two periodograms given in Figure 2. One of these is the correct periodogram for the logarithm of the UK gas data while the other is the periodogram for some other dataset. Identify the correct periodogram for the logarithm of the UK gas data giving reasons for your answer.

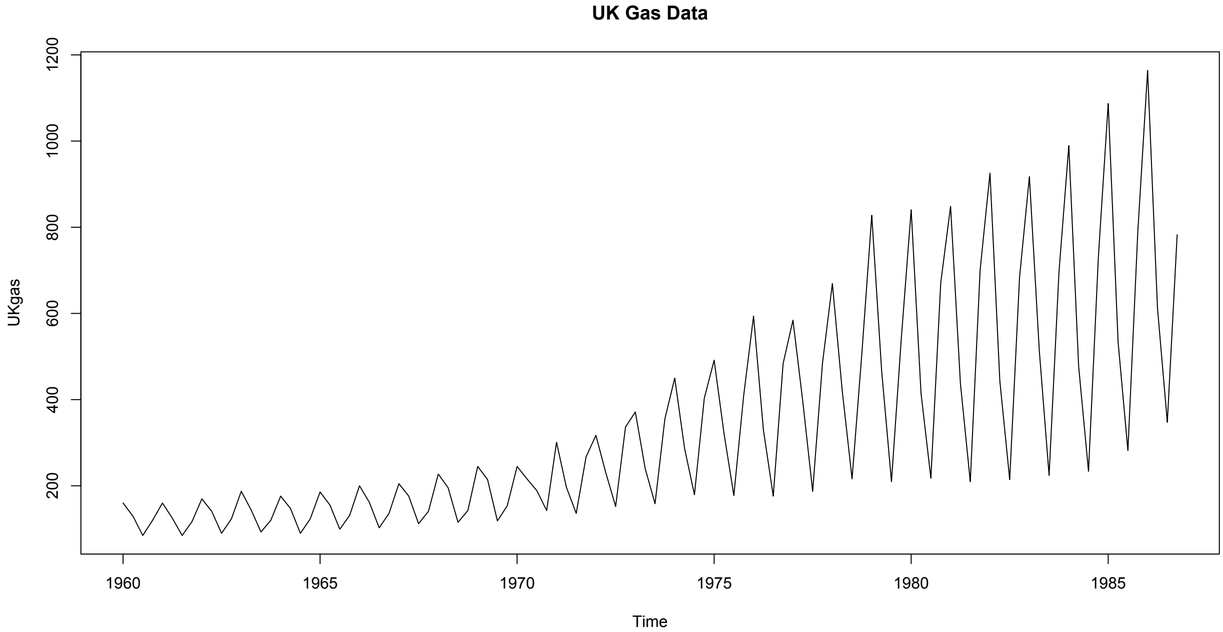

Figure 1: UKgas Data

Solution: The first periodogram is the correct one for the logarithm of the UK gas data. The UK gas data is quarterly so it is natural to expect sinusoids with a period of 4, or frequency of $1/4 = 0.25$. There is a spike in the first periodogram at 0.25. Further, there is also a spike in the first periodogram at the first Fourier frequency $(1/n)$. This sinusoid is there to capture the trend that is present in the dataset. In contrast, both these features are absent in the second periodogram.

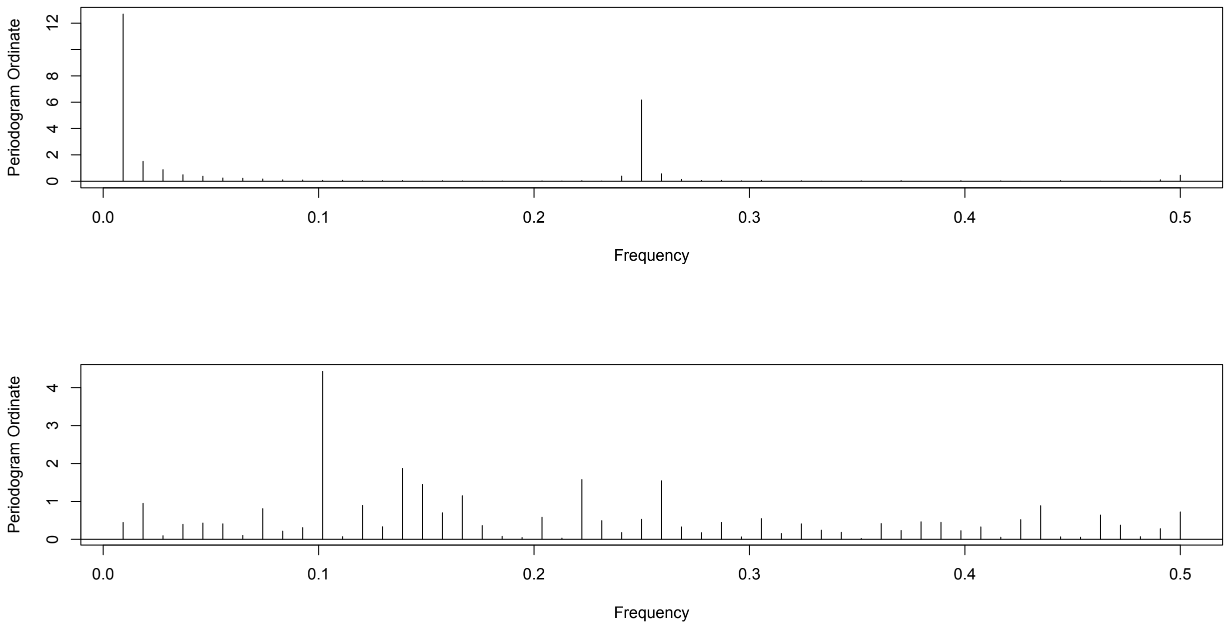

Figure 2: Two Periodograms

3. For a time series dataset $y_1, \dots, y_n$, I would like to fit the model:

$$y_t = \beta_0 + [\beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)] \exp(-\omega t) + \epsilon_t \qquad \text{with } \epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2).$$

This fits a sinusoid to the data with an exponential decay. The model has six unknown parameters $\beta_0, \beta_1, \beta_2, f, \omega, \sigma$. Suppose my main interest is in the parameters f and ω. Describe a procedure for estimating f and ω along with proper uncertainty quantification.

Solution: We can rewrite this model as

$$Y = X_{f,\omega}\beta + \epsilon$$

where Y is the $n \times 1$ vector with components $Y_1, \dots, Y_n$, and $X_{f,\omega}$ is a $n \times 3$ matrix whose first column has all ones, second column has the entries $\cos(2\pi f t) \exp(-\omega t)$ for $t = 1, \dots, n$ and third column has the entries $\sin(2\pi f t) \exp(-\omega t)$ for $t = 1, \dots, n$. Further β has the three components $\beta_0, \beta_1, \beta_2$ and ϵ has the components $\epsilon_1, \dots, \epsilon_n$.

We studied inference for parameters for models similar to this in class. In particular, taking a uniform Unif$(-C, C)$ prior on $\beta_0, \beta_1, \beta_2, \log\sigma$, we derived that the (unnormalized) posterior for f, ω is given by

$$|X'_{f,\omega} X_{f,\omega}|^{-1/2} \left(\frac{1}{RSS(f,\omega)}\right)^{(n-3)/2} \tag{4}$$

Here $RSS(f, \omega)$ is defined as

$$RSS(f,\omega) = \min_{\beta_0,\beta_1,\beta_2} \sum_{t=1}^{n} \left(y_t - \beta_0 - [\beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)] \exp(-\omega t)\right)^2$$

Using the above posterior, we can do inference on f and ω via the following numerical procedure:

a) Take a grid of possible values of f and ω. For example, we can take f to be in a dense grid in the range 0.01 to 0.5. ω will be in a dense grid in some large range $(-C, C)$ (or just $(0, C)$ if we believe ω to be a positive decay parameter for the exponential).

b) For each possible value of f and ω, calculate the value of the unnormalized posterior (4). We then normalize these values so they sum to one. This discrete distribution approximates the posterior of (f, ω).

c) We can marginalize to obtain discrete approximations for the separate posterior distributions of f and ω. These can be used to obtain estimates and uncertainty quantification for f and ω (estimate is the posterior mean and uncertainty can be quantified by the posterior standard deviation).

4. The data plotted in Figure 3 gives (seasonally adjusted) monthly observations on the retail sales (in millions of dollars) of Furniture Stores. For each of the following models, indicate whether they are adequate for this dataset giving reasons:

   - **Model One**: $y_t = \beta_0 + \beta_1(t-1) + \epsilon_t$ with $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. This model has three parameters β_0, β_1 and σ.

   - **Model Two**: $y_t = \beta_0 + \beta_1(t-1) + \beta_2(t-\omega)_+ + \epsilon_t$ with $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. This model has five parameters $\beta_0, \beta_1, \beta_2, \omega, \sigma$. Here x_+ denotes $\max(x, 0)$.

   - **Model Three**: $y_t = \beta_0 + \beta_1(t-1) + \beta_2(t-\omega_1)_+ + \beta_3(t-\omega_2)_+ + \epsilon_t$ with $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. This model has seven parameters $\beta_0, \beta_1, \beta_2, \beta_3, \omega_1, \omega_2, \sigma$.

   - **Model Four**: $y_t = \beta_0 + \beta_1(t-1) + \sum_{j=2}^{n-1} \beta_j(t-j)_+ + \epsilon_t$ with $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$.

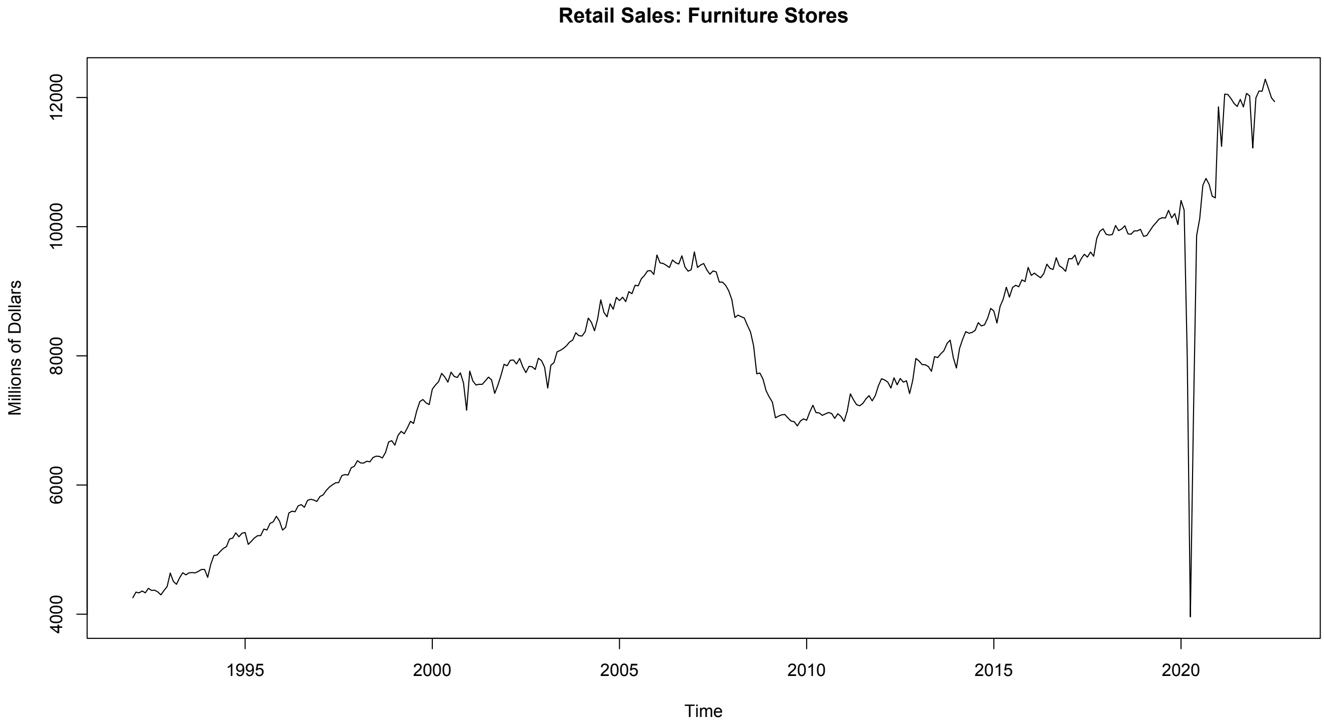

Figure 3: FRED Data

Solution: Here are the comments on the appropriateness of each of the three models for the given dataset:

a) **Model One**: This model just fits a simple linear trend which is clearly not appropriate. The data shows an increasing linear trend until about 2006 but from 2006 to about 2010, there is a decreasing linear trend, after which there is again an increasing linear trend (except around Covid). Such behavior will not be explained by this simple linear trend model.

b) **Model Two**: This model fits a linear trend until time ω and then another linear trend (with different slope) after time ω. As just described, the data shows at least three separate linear trends so this model also is not appropriate.

c) **Model Three**: This model can explain three separate linear trends so it will work much better than the previous two models. However, even this model is not fully appropriate. It will ignore nonlinear behaviour in certain parts, and might substantially smooth out the dip in sales during COVID time.

d) **Model Four**: This is a high-dimensional model having n coefficient parameters (as well as σ). It will overfit the data if no regularization is imposed on the coefficient parameters. With proper regularization (e.g., $\sum_{j=2}^{n-1} \beta_j^2$ or $\sum_{j=2}^{n-1} |\beta_j|$), this model can be useful and fit a smooth trend function to the data.

5. Consider the time series dataset plotted in Figure 4. This is a monthly dataset with 271 observations one for each month from January 2000 to July 2022. The dataset is collected from FRED and gives the total Natural Gas Consumption in the United States.

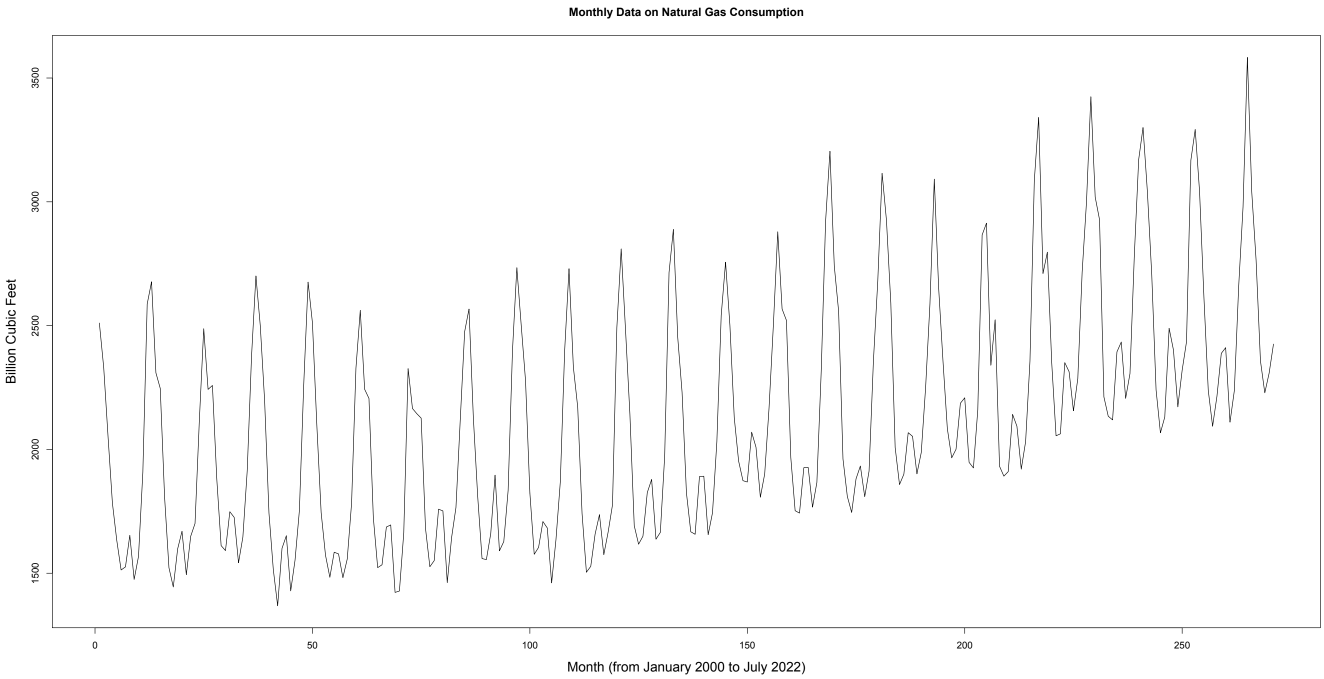

Figure 4: Natural Gas Consumption Data

Consider the two periodograms given in Figure 5. One of these is the correct periodogram for the data in Figure 4 while the other is the periodogram for some other dataset. Identify the correct periodogram giving reasons for your answer.

Solution: The dataset clearly exhibits a strong annual (12-month) cycle (because heating demand peaks in winter and typically falls off in summer). Therefore we would expect a large peak at the annual frequency $1/12$ in its periodogram. We may also see smaller peaks at frequencies $2/12, 3/12$ reflecting half-year or quarterly subcycles. Further, at very low frequencies (far left of the periodogram), we would expect to see some peaks corresponding to the increasing trend in the data. All these features in the second periodogram (and none of these features are present in the first periodogram). Thus the correct periodogram for the Gas Consumption data is the second periodogram in Figure 5.

6. An observed time series dataset is plotted in Figure 6 along with its periodogram and the logarithm of the periodogram. To this dataset, we use the model:

$$I(j/n) \overset{\text{ind}}{\sim} \frac{\gamma_j^2}{n}\chi_2^2 \qquad \text{for } j = 1, \dots, m$$

where $m = \lfloor n/2 \rfloor$. We also assume that $\alpha_j = \log \gamma_j$ is smooth in j, and then use the following three methods for estimating $\alpha_j, 1 \le j \le m$:

a) **Method One**: Minimize $\sum_{j=1}^{m}\left(\frac{nI(j/n)}{2}e^{-2\alpha_j} + 2\alpha_j\right) + \lambda\sum_{j=2}^{m-1}(\alpha_{j+1} - 2\alpha_j + \alpha_{j-1})^2$ for $\lambda = 100$

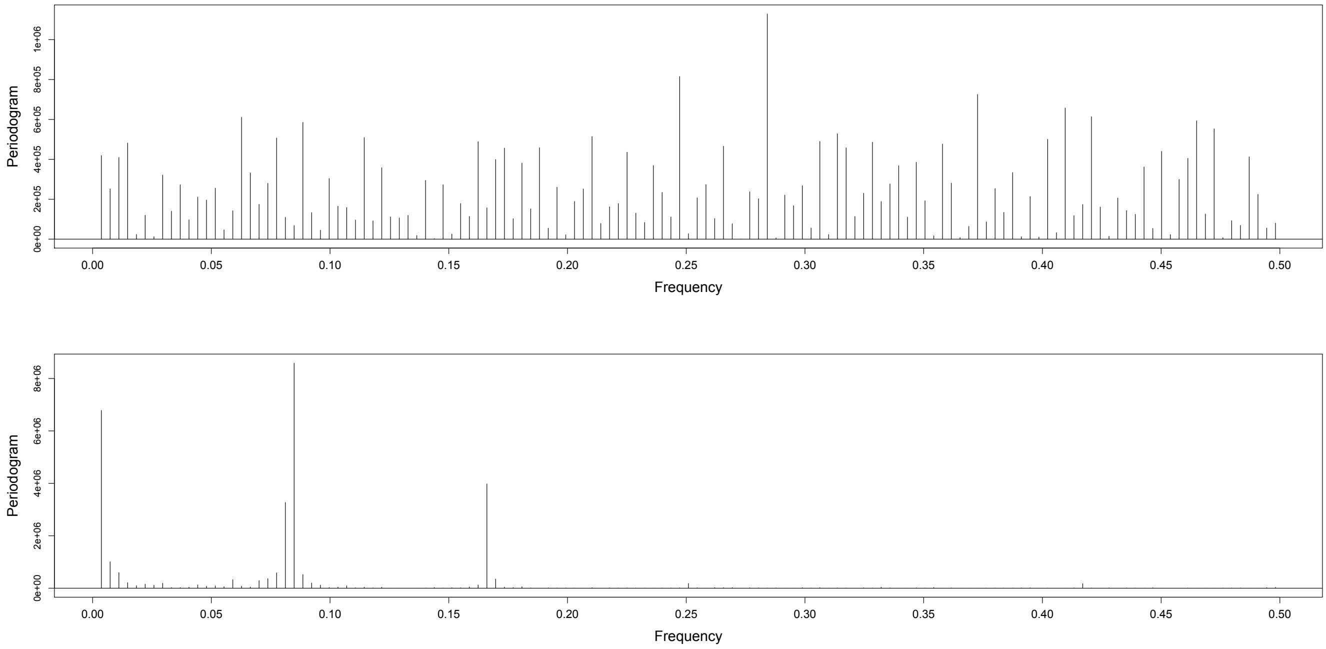

Figure 5: Two Periodograms

b) **Method Two**: Minimize $\sum_{j=1}^{m}\left(\frac{nI(j/n)}{2}e^{-2\alpha_j}+2\alpha_j\right)+\lambda\sum_{j=2}^{m-1}(\alpha_{j+1}-2\alpha_j+\alpha_{j-1})^2$ for $\lambda=10000$

c) **Method Three**: Minimize $\sum_{j=1}^{m}\left(\frac{nI(j/n)}{2}e^{-2\alpha_j}+2\alpha_j\right)+\lambda\sum_{j=2}^{m-1}|\alpha_{j+1}-2\alpha_j+\alpha_{j-1}|$ for $\lambda=100$.

In Figure 7, we plot log periodogram $\log I(j/n)$ along with the result of the above three methods: $\log(2\hat{\gamma}_j^2/n)$ (where $\hat{\gamma}_j=\exp(\hat{\alpha}_j)$). Match the methods with the correct plots in Figure 7 giving reasons.

Solution: From Figure 7, the middle estimate (Estimator Two) is piecewise linear, while the other two estimates are not piecewise linear. Since L_1 penalty leads to sparse solutions, the middle subplot (Estimator Two) should correspond to the LASSO penalty (Method Three).

The first subplot in Figure 7 is much smoother compared to the third subplot. Therefore the λ parameter should be larger for the first subplot. Thus, Method One corresponds to Estimator Three, and Method Two corresponds to Estimator One.

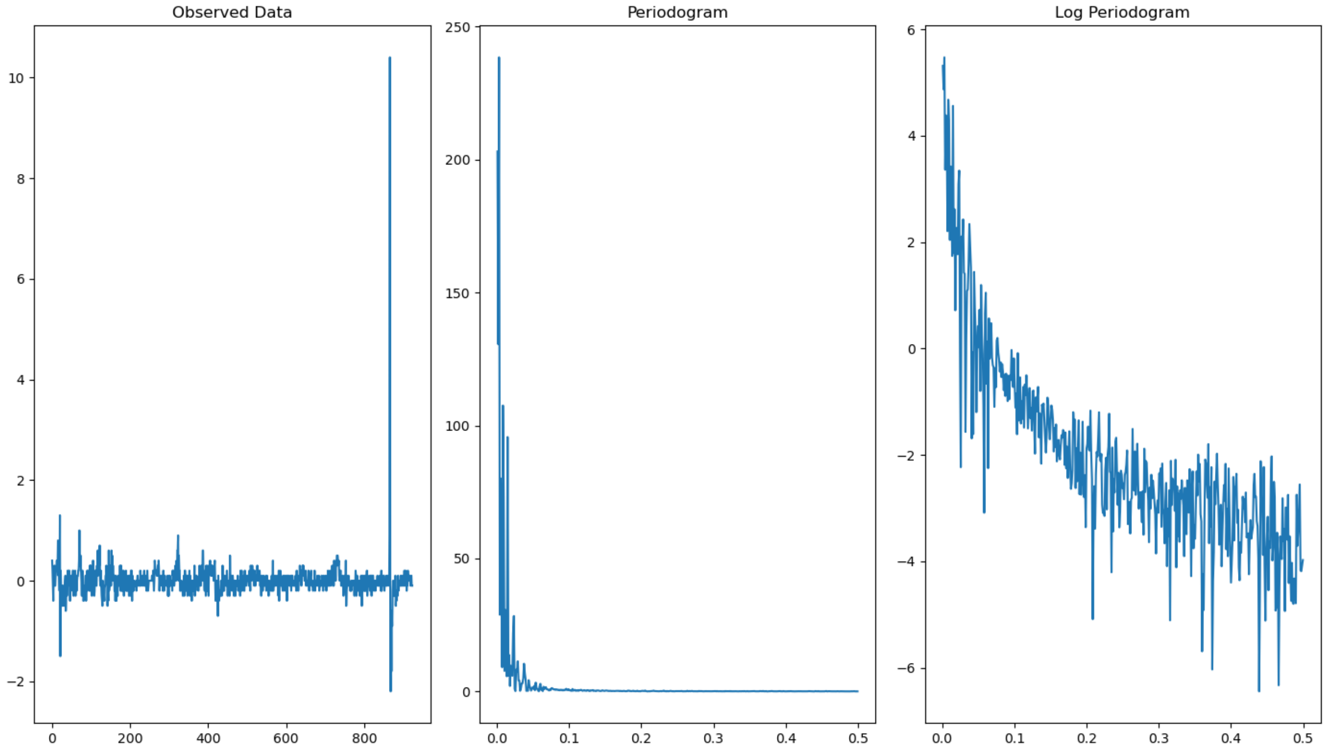

Figure 6: Observed data, periodogram and log periodogram

7. Figure 8 contains the plot of a time series $y_1, \dots, y_n$ (with $n = 2048$) along with three trend estimates. Each of the trend estimates correspond to one of the following minimization problems. Identify the correct ones.

   a) $\sum_{t=1}^{n}(y_t - \mu_t)^2 + \lambda \sum_t |\mu_t|$.

   b) $\sum_{t=1}^{n}(y_t - \mu_t)^2 + \lambda \sum_{t=2}^{n-1} |\mu_{t+1} - 2\mu_t + \mu_{t-1}|$.

   c) $\sum_{t=1}^{n}(y_t - \mu_t)^2 + \lambda \sum_{t=1}^{n-1} |\mu_{t+1} - \mu_t|$.

   d) $\sum_{t=1}^{n}(y_t - \mu_t)^2 + \lambda \sum_{t=1}^{n-1}(\mu_{t+1} - \mu_t)^2$.

   Solution: Trend estimate one is piecewise constant, which means that $\mu_{t+1} - \mu_t$ is sparse. So it should correspond to the minimization problem (c) because the L_1 penalty on $\mu_{t+1} - \mu_t$ leads to sparsity of $\mu_{t+1} - \mu_t$.

   Trend estimate two is piecewise linear, which means that $(\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1}) = \mu_{t+1} - 2\mu_t + \mu_{t-1}$ is sparse. So it should correspond to the minimization problem (c) because the L_1 penalty on $\mu_{t+1} - 2\mu_t + \mu_{t-1}$ leads to sparsity of $\mu_{t+1} - 2\mu_t + \mu_{t-1}$.

   Trend estimate three is quite smooth. Of the given options, the one which is likely to result in this estimate is (d). If λ is chosen large, then $\mu_{t+1} - \mu_t$'s will be small leading to the smooth appearance.

   Note that (a) will lead to sparse estimates of $\{\mu_t\}$ i.e., most of the μ_t's will equal zero exactly. This is not appropriate for this dataset and does not correspond to any of the three trend estimates.

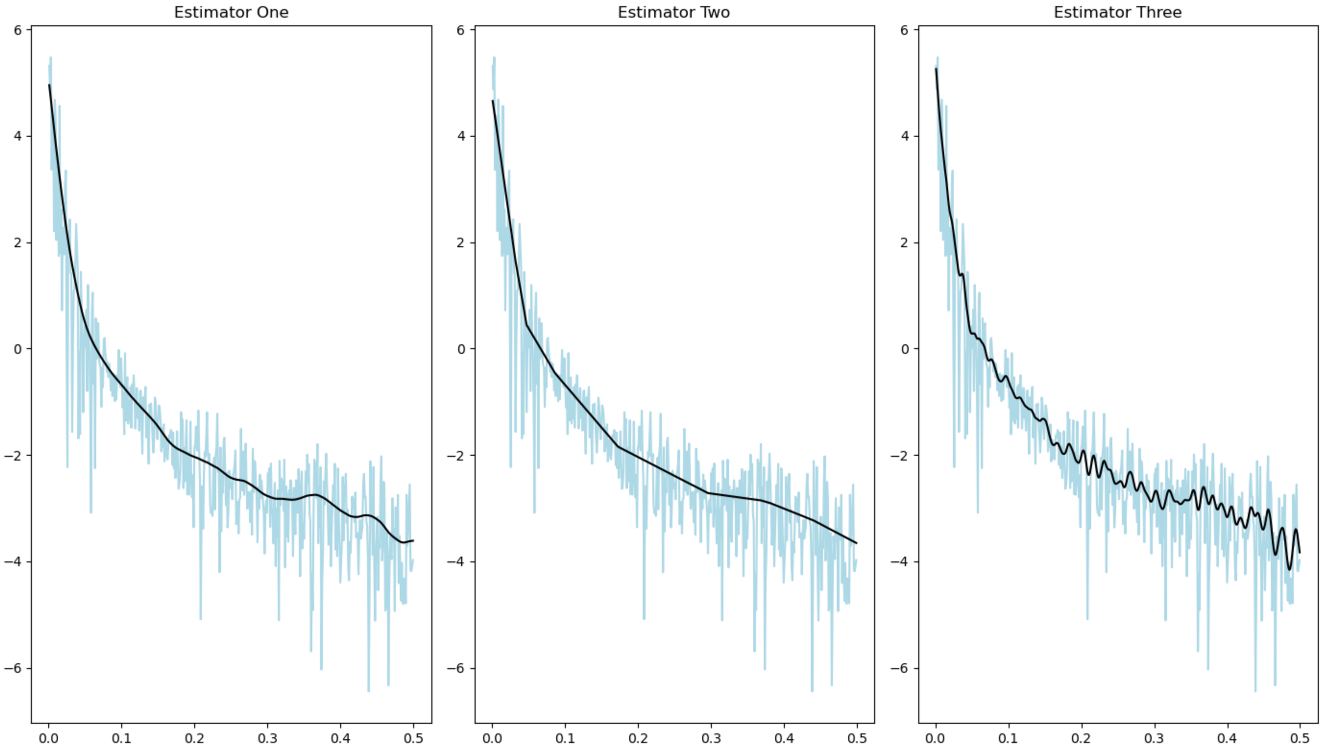

Figure 7: Three log spectrum estimates

8. Given a observed time series $y_1, \dots, y_n$, I want to fit the model:

$$y_t \overset{\text{ind}}{\sim} N(0, \exp(2(\beta_0 + \beta_1 t))).$$

   a) Write the log-likelihood for the parameters β_0, β_1.

   b) Let $\hat{\beta}_0$ and $\hat{\beta}_1$ denote the MLEs of β_0 and β_1 respectively. Show that

$$\hat{\beta}_0 = \frac{1}{2} \log \left(\sum_{t=1}^{n} y_t^2 e^{-2t\hat{\beta}_1} \right) - \frac{1}{2} \log n. \tag{5}$$

Solution: This model can be written as: $y_t \sim N(0, \tau_t^2)$ with $\log \tau_t = \beta_0 + \beta_1 t$. The likelihood (similar to Model Two in Lecture 14) is proportional to

$$\prod_{t=1}^{n} \frac{1}{\tau_t} \exp \left(-\frac{y_t^2}{2\tau_t^2} \right)$$

so that the log-likelihood is

$$\sum_{t=1}^{n} \left(-\log \tau_t - \frac{y_t^2}{2\tau_t^2} \right) \qquad \text{with } \tau_t = \exp(\beta_0 + \beta_1 t).$$

Plugging in $\log \tau_t = \beta_0 + \beta_1 t$, we get the following formula for the log-likelihood:

$$\sum_{t=1}^{n} \left(-\beta_0 - \beta_1 t - \frac{y_t^2}{2} \exp(-2\beta_0 - 2\beta_1 t) \right)$$

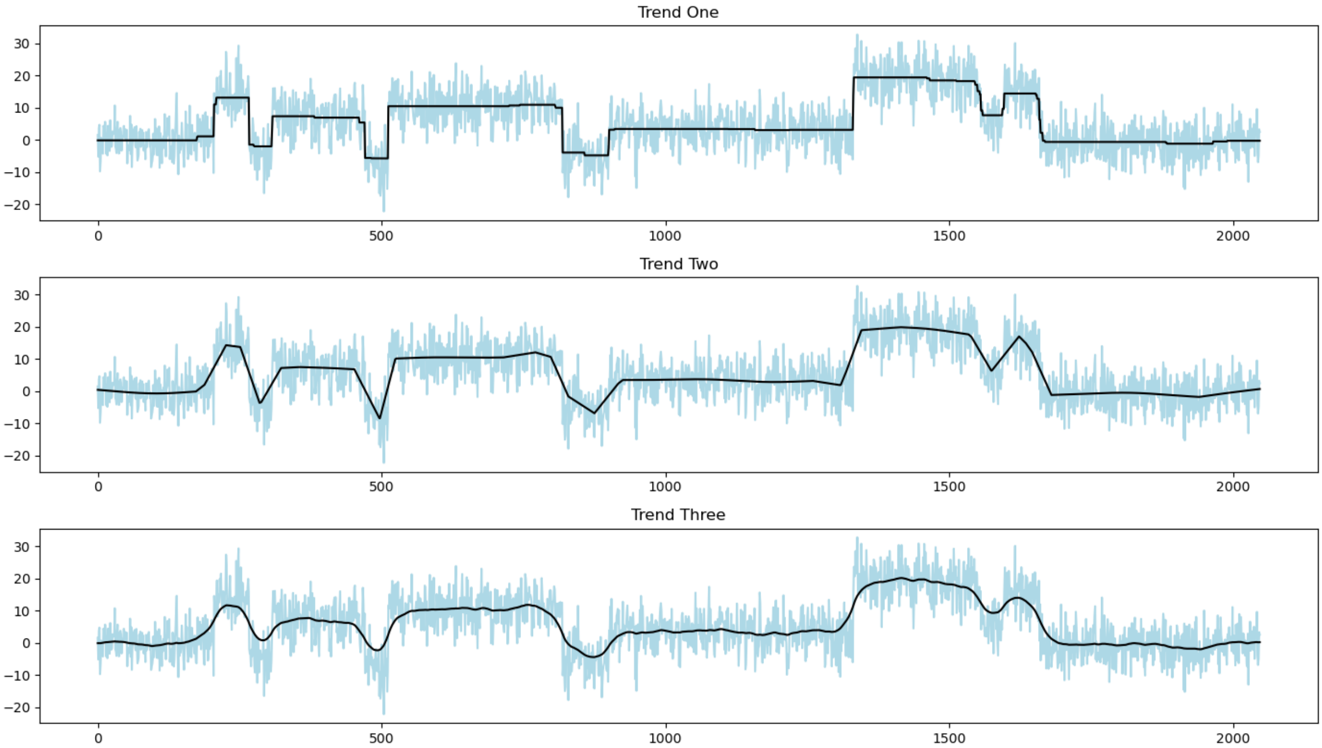

Figure 8: Three Trend estimates

We need to take derivatives w.r.t β_0, β_1 and set them to zero to get the MLEs. The derivative w.r.t β_0 is

$$-n + \sum_{t=1}^{n} y_t^2 \exp(-2\beta_0 - 2\beta_1 t).$$

Setting it to zero and solving for β_0, we get

$$\exp(-2\beta_0) = \frac{n}{\sum_{t=1}^{n} y_t^2 \exp(-2\beta_1 t)}$$

which is equivalent to (5).