STAT 153 - Introduction to Time Series
Practice Midterm

Fall 2022, UC Berkeley

1. 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.

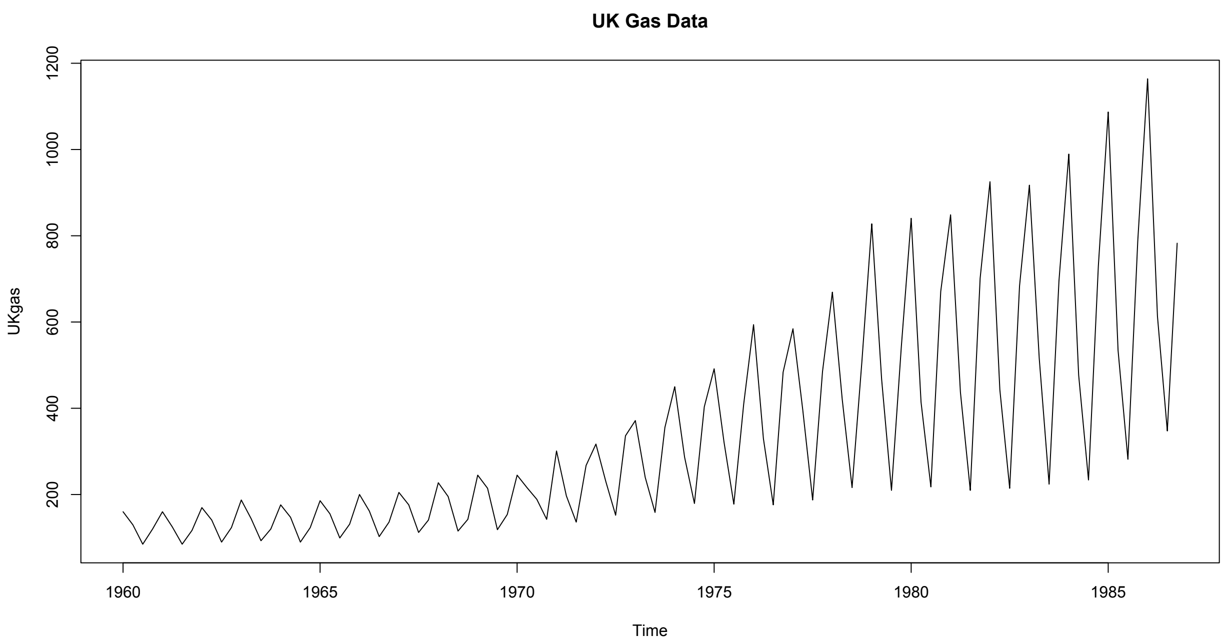

Figure 1: UKgas Data

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.

2. 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:

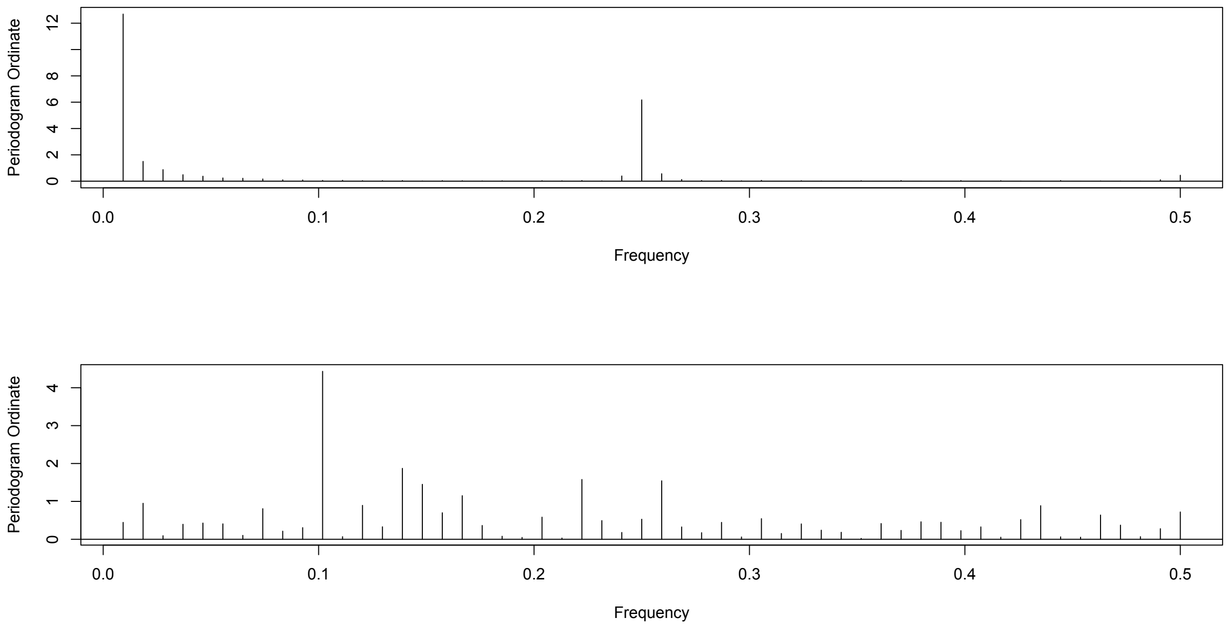

Figure 2: Two Periodograms

- **Model One**: $Y_t = \beta_0 + \beta_1 t + Z_t$ with $Z_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. This model has three parameters β_0, β_1 and σ.
- **Model Two**: $Y_t = \beta_0 + \beta_1 t + \beta_2 (t - \omega)_+ + Z_t$ with $Z_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. This model has five parameters $\beta_0, \beta_1, \beta_2, \omega, \sigma$. Here x_+ denotes $\max(x, 0)$.
- **Model Three**: $Y_t = \beta_0 + \beta_1 t + \beta_2 (t - \omega_1)_+ + \beta_3 (t - \omega_2)_+ + Z_t$ with $Z_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. This model has seven parameters $\beta_0, \beta_1, \beta_2, \beta_3, \omega_1, \omega_2, \sigma$.

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. For example, it will, most likely, not be able to explain the sudden dip in sales during Covid time.

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

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

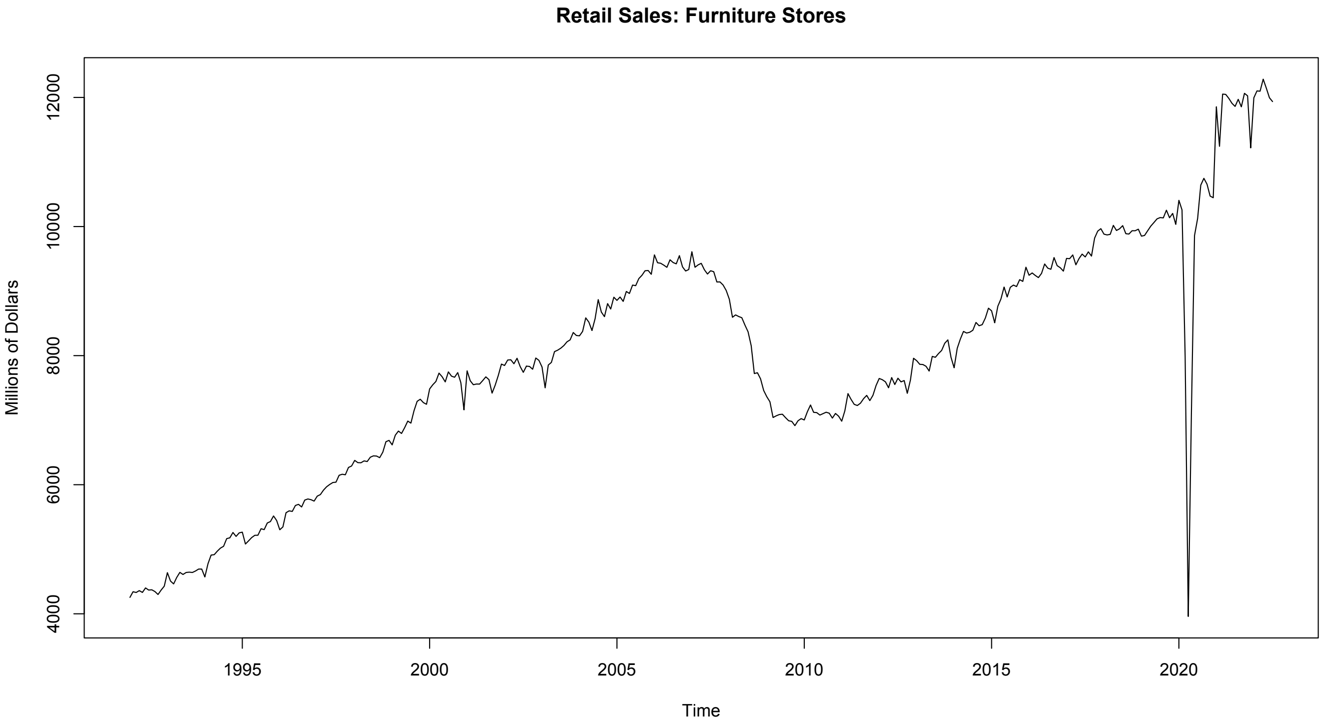

Figure 3: FRED Data

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

$$Y = X(f, \omega)\beta + Z$$

where Y is the $n \times 1$ vector with components $Y_1, \ldots, Y_n$, 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 $t = 1, \ldots, n$ and third column has the entries $\sin(2\pi f t)\exp(-\omega t)$ for $t = 1, \ldots, n$. Further β has the three components $\beta_0, \beta_1, \beta_2$ and Z has the components $Z_1, \ldots, Z_n$. We studied inference for parameters for this model in class. In particular, taking a uniform Unif$(-C, C)$ prior on $f, \omega, \beta_0, \beta_1, \beta_2, \log \sigma$, we derived that the (unnormalized) posterior for f, ω is given by

$$|X(f, \omega)' X(f, \omega)|^{-1/2} \|Y - X(f, \omega)\hat{\beta}(f, \omega)\|^{-(n-p)} \tag{1}$$

with $p = 2$. Here $\hat{\beta}(f, \omega) = (X'X)^{-1}X'Y$ with $X = X(f, \omega)$. 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 (1). 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. For a time series dataset $y_1, \dots, y_n$, I would like to fit the model:

$$Y_t = \beta_0 + \beta_1 t + Z_t \qquad \text{with } Z_t \overset{\text{i.i.d}}{\sim} C(0, \sigma).$$

where $C(0, \sigma)$ is the Cauchy density with scale σ (its density is $x \mapsto \frac{\sigma}{\pi(\sigma^2+x^2)}$). This model has the three unknown parmaeters β_0, β_1, σ. Describe a numerical procedure for estimating β_1 and quantifying uncertainty.

Solution: We shall take the following prior on β_0, β_1, σ:

$$\beta_0, \beta_1, \log\sigma \overset{\text{i.i.d}}{\sim} \text{Unif}(-C, C)$$

for a large constant C. The posterior is then given by

$$\begin{aligned} f_{\beta_0,\beta_1,\sigma|\text{data}}(\beta_0, \beta_1, \sigma) &\propto \frac{I\{-C < \beta_0, \beta_1, \log\sigma < C\}}{(2C)^3\sigma} \prod_{t=1}^{n} \left(\frac{1}{\pi} \frac{\sigma}{\sigma^2 + (y_t - \beta_0 - \beta_1 t)^2} \right) \\ &\propto \frac{I\{-C < \beta_0, \beta_1, \log\sigma < C\}}{\sigma} \prod_{t=1}^{n} \left(\frac{\sigma}{\sigma^2 + (y_t - \beta_0 - \beta_1 t)^2} \right). \end{aligned}$$

From the above unnormalized posterior, inference for the parameters β_0, β_1, σ can be done via the following numerical procedure:

a) Take a grid of possible values of $\beta_0, \beta_1, \log\sigma$ in $(-C, C)$.

b) For each possible value of $\beta_0, \beta_1, \log\sigma$ in the grid, calculate the value of the unnormalized posterior given above. We then normalize these values so they sum to one. This discrete distribution approximates the posterior of $(\beta_0, \beta_1, \sigma)$.

c) We can marginalize to obtain discrete approximations for the separate posterior distributions of β_0, β_1, σ. These can be used to obtain estimates and uncertainty quantification for the parameters separately (estimate is the posterior mean and uncertainty can be quantified by the posterior standard deviation).

5. I have an observed time series dataset $y_1, \dots, y_n$ with $n = 100$. For this dataset, I am considering the two models:

- **Model One**: $y_1, \dots, y_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma^2)$. This model has two unknown parameters μ and σ^2.
- **Model Two**: $y_1, \dots, y_n$ are independent with

$$y_1, \dots, y_{n/2} \overset{\text{i.i.d}}{\sim} N(\mu_1, \sigma^2) \quad \text{and} \quad y_{(n/2)+1}, \dots, y_n \overset{\text{i.i.d}}{\sim} N(\mu_2, \sigma^2).$$

This model has three unknown parameters μ_1, μ_2 and σ^2.

a) Describe a Bayesian model selection procedure for deciding between the above two models.

b) Write down the AIC of each of the two models explicitly in terms of $y_1, \dots, y_n$.

Solution: Both Model One and Model Two are regression models:

$$Y = X\beta + Z$$

for appropriate X. More precisely, Model One corresponds to X having just the single column of all ones ($p = 1$). Model Two corresponds to X having two columns ($p = 2$): the first column has all ones and the second column has zeros in the first $n/2$ rows and ones in the remaining rows. In class, we saw that Bayesian model selection for linear models can be done via the following formula:

$$\text{Evidence}(M) \propto \frac{\Gamma(p/2)}{\|X\hat{\beta}\|^p} \frac{\Gamma(\frac{n-p-1}{2})}{\|Y - X\hat{\beta}\|^{n-p}}.$$

For Model One, $p = 1$ and it is easy to see that $X\hat{\beta}$ is the $n \times 1$ vector all of whose entries are equal to $\bar{y} := (y_1 + \dots + y_n)/n$. Thus

$$\|X\hat{\beta}\|^p = \left(\sqrt{\bar{y}^2 + \dots + \bar{y}^2}\right)^p = n^{p/2}|\bar{y}|^p = \sqrt{n}|\bar{y}|$$

and

$$\|Y - X\hat{\beta}\|^{n-p} = \left(\sum_{i=1}^{n} (y_i - \bar{y})^2\right)^{(n-p)/2} = \left(\sum_{i=1}^{n} (y_i - \bar{y})^2\right)^{(n-1)/2}$$

Thus

$$\text{Evidence}M_1) \propto \frac{\Gamma(1/2)}{\sqrt{n}|\bar{y}|} \frac{\Gamma\left(\frac{n-2}{2}\right)}{\left(\sum_{i=1}^{n} (y_i - \bar{y})^2\right)^{(n-1)/2}} \tag{2}$$

For Model Two, $p = 2$ and $X\hat{\beta}$ is the $n \times 1$ vector whose first $n/2$ entries are equal to $\bar{y}^{(1)}$ and the last $n/2$ entries are equal to $\bar{y}^{(2)}$ where

$$\bar{y}^{(1)} := \frac{y_1 + \dots + y_{n/2}}{n/2} \quad \text{and} \quad \bar{y}^{(2)} := \frac{y_{(n/2)+1} + \dots + y_n}{n/2}.$$

Thus

$$\|X\hat{\beta}\|^p = \frac{n}{2}\left(\bar{y}^{(1)}\right)^2 + \frac{n}{2}\left(\bar{y}^{(2)}\right)^2$$

and

$$\|Y - X\hat{\beta}\|^{n-p} = \left[\sum_{i=1}^{n/2} \left(y_i - \bar{y}^{(1)}\right)^2 + \sum_{i=(n/2)+1}^{n} \left(y_i - \bar{y}^{(2)}\right)^2\right]^{(n-2)/2}.$$

Therefore

$$\text{Evidence}(M_2) \propto \frac{\Gamma(1)}{\frac{n}{2}\left(\bar{y}^{(1)}\right)^2 + \frac{n}{2}\left(\bar{y}^{(2)}\right)^2} \frac{\Gamma\left(\frac{n-3}{2}\right)}{\left[\sum_{i=1}^{n/2} \left(y_i - \bar{y}^{(1)}\right)^2 + \sum_{i=(n/2)+1}^{n} \left(y_i - \bar{y}^{(2)}\right)^2\right]^{(n-2)/2}} \tag{3}$$

We calculate the right hand sides of (2) and (3); and then normalize them so they sum to 1. These give posterior probabilities for the two models and we choose the model with the higher posteriorr probability.

The AIC for a linear model was calculated in Lecture 10 to be

$$AIC(M) = n + n\log\left(\frac{2\pi}{n}\|Y - X\hat{\beta}\|^2\right) + 2(p+1)$$

Thus for model one, we get

$$AIC(M_1) = n + n \log \left(\frac{2\pi}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2 \right) + 4.$$

For model two,

$$AIC(M_2) = n + n \log \left(\frac{2\pi}{n} \left[\sum_{i=1}^{n/2} \left(y_i - \bar{y}^{(1)} \right)^2 + \sum_{i=(n/2)+1}^{n} \left(y_i - \bar{y}^{(2)} \right)^2 \right] \right) + 6.$$