Time Series Analysis: Interesting Problems

Jonathan Mostovoy
University of Toronto

March 31, 2017

Contents

1 Theory

1.1 The Discrete Fourier Transform

Let $\{x_t\}$ and $\{y_t\}$ be two infinite sequences of real-numbers and define:

$$X(\omega) = \sum_{t=-\infty}^{\infty} x_t e^{2\pi i\omega t} \quad \text{and} \quad Y(\omega) = \sum_{t=-\infty}^{\infty} y_t e^{2\pi i\omega t}$$

to be their (discrete) Fourier transforms. Define $\{z_t\}$ to be the convolution of $\{x_t\}$ and $\{y_t\}$; i.e.:

$$z_t = \sum_{u=-\infty}^{\infty} x_u y_{t-u}$$

(a) *If $Z(\omega)$ is the Fourier transform of $\{z_t\}$, show that:*

$$Z(\omega) = X(\omega)Y(\omega)$$

(b) *Suppose that $\{x_t\}$ is a time series and $\{y_t\}$ is defined by:*

$$y_t = \sum_{u=0}^{p} c_u x_{t-u}$$

Then from (a), it follows that the periodograms of $\{y_t\}$ and $\{x_t\}$ are related by:

$$I_y(\omega) \approx |\Gamma(\omega)|^2 I_x(\omega)$$

Where $\Gamma(\omega) = \sum_{u=0}^{p} c_u e^{2\pi i\omega u}$. How does this explain the behaviour of the periodograms in parts (a) and (b) of Question 1? (Hint: Take $c_0 = 1$, $c_1 = -1$.)

(c) *Suppose that $c_u > 0$ for $u = 0, \dots, p$ with $\sum_{j=0}^{p} c_j = 1$. How does the periodogram of $\{y_t\}$ (defined in (b)) compare to that of $\{x_t\}$?*

(a) *Proof.* We prove this fact directly:

$$\begin{aligned}
X(\omega)Y(\omega) &= \sum_{u=-\infty}^{\infty} x_u e^{2\pi i\omega u} \sum_{v=-\infty}^{\infty} y_v e^{2\pi i\omega v} \\
&= \sum_{v=-\infty}^{\infty} \sum_{u=-\infty}^{\infty} x_u y_v e^{2\pi i\omega(u+v)} \\
&= \sum_{t=-\infty}^{\infty} \left(\sum_{u=-\infty}^{\infty} x_u y_{t-u} \right) e^{2\pi i\omega(u+(t-u))} && \text{If we let } t = v + u \\
&= \sum_{t=-\infty}^{\infty} z_t e^{2\pi i\omega t} \\
&= Z(\omega)
\end{aligned}$$

□

(b) *Proof.* If we do as suggested (take $c_0 = 1, c_1 = -1, p = 1$) we will have:

$$\Gamma(\omega) = e^0 - e^{2\pi i\omega} = 1 - e^{2\pi i\omega}$$

As such,

$$\begin{aligned}
|\Gamma(\omega)|^2 &= \Gamma(\omega)\overline{\Gamma(\omega)} \\
&= \left(1 - e^{2\pi i\omega}\right)\left(1 - \overline{e^{2\pi i\omega}}\right) \\
&= \left(1 - e^{2\pi i\omega}\right)\left(1 - e^{-2\pi i\omega}\right) \\
&= 2 - (2\mathrm{Re}(e^{2\pi i\omega})) \\
&= 2 - 2\cos(2\pi\omega) \\
&= 4\sin^2(\pi\omega)
\end{aligned}$$

Therefore, taking the first differences removes low frequency variation while emphasised high frequency variation. In problem 1, taking first differences removed the peak at $\omega = 0$ in the periodogram.

□

(c) *Proof.* We recall that the inverse Fourier transform is defined as:

$$\mathcal{F}^{-1}(U(\omega)) := \sum_{\omega=-\infty}^{\infty} U(\omega)e^{-2\pi i\omega t} = u_t$$

As such, we see that $\mathcal{F}^{-1}(\Gamma(\omega)) = c_u$, and hence we may derive the functional relationship:

$$y_t = \mathcal{F}^{-1}(\Gamma(\omega))x_{t-u} = \mathcal{F}(\Gamma(\omega) * X(\omega))$$

Where $*$ denotes the discrete convolution. Intuitively, this makes sense because we are weighting y_t on the past histories of x_t though a Fourier transform which permits us to operate between frequencies and time. □

1.2 Complex Trigonometric Functions

(a) *Show that:*

$$\sum_{k=0}^{n-1} e^{ik\theta} = \sum_{k=0}^{n-1} \left[e^{i\theta}\right]^k = \frac{1 - e^{in\theta}}{1 - e^{i\theta}}$$

if θ is not a multiple of 2π and use this result to give simple expressions for:

$$\sum_{k=0}^{n-1} \cos(k\theta) \quad \text{and} \quad \sum_{k=0}^{n-1} \sin(k\theta)$$

(b) *Let $x_1, \dots, x_n$ be a sequence of numbers (for example, a time series) and define the discrete Fourier transform as:*

$$X(\omega) = \sum_{t=1}^{n} x_t e^{2\pi i \omega t}$$

Define $\omega_k = \frac{k}{n}$ for $k = 0, \dots, n-1$. Show that:

$$x_s = \frac{1}{n} \sum_{k=0}^{n-1} X(\omega_k) e^{-2\pi i \omega_k s}$$

(a) *Proof.* For the first equality $\left(\sum_{k=0}^{n-1} e^{ik\theta} = \sum_{k=0}^{n-1} \left[e^{i\theta}\right]^k\right)$, we recall the following from an elementary course in Complex Analysis:

(a) $e^{ix} = \cos(x) + i\sin(x)$, where $x \in \mathbb{R}$. (Euler's Formula)
(b) $\big(\cos(x) + i\sin(x)\big)^n = \cos(nx) + i\sin(nx)$, where $x \in \mathbb{R}, n \in \mathbb{Z}$. (De Moivre's Formula)
(c) $\cos(\alpha z) := \frac{1}{2}\left(e^{i\alpha z} + e^{-i\alpha z}\right)$ and $\sin(\alpha z) := \frac{1}{2i}\left(e^{i\alpha z} - e^{-i\alpha z}\right)$

The first two formulae imply the first equality.

For the second equality, we consider: an arbitrary summation defined by $\sum_{i=0}^{m} w^i$, $w \in \mathbb{C}$. Then we will have:

$$w\sum_{i=0}^{m} w^i = \sum_{i=0}^{m} w^i + q^{k^{m+1}} - 1 \implies \text{ if } q \neq 1, \sum_{i=0}^{m} = \frac{1-q^{m+1}}{1-q} \tag{1}$$

Next, we note that $e^{i\theta k} \neq 1 \iff \theta k \neq 2\pi k, k \in \mathbb{Z}$. As such, $\forall\ \theta \neq 2\pi$:

$$\sum_{k=0}^{n-1} \left(e^{i\theta}\right)^k = \frac{1-\left(e^{i\theta}\right)^{n-1+1}}{1-e^{i\theta}} = \frac{1-e^{i\theta n}}{1-e^{i\theta}} \qquad \text{By (1)}$$

Next, we notice the following:

$$\sum_{k=0}^{n-1} \left(e^{i\theta}\right)^k = 1 + \cos\big(\theta\big) + i\sin\big(\theta\big) + \cos\big(2\theta\big) + i\sin\big(2\theta\big) + \cdots + \cos\big((n-1)\theta\big) + i\sin\big((n-1)\theta\big)$$

$$\implies \text{Re}\left(\sum_{k=0}^{n-1} \left(e^{i\theta}\right)^k\right) = \sum_{k=0}^{n-1} \cos(k\theta) \quad \text{and} \quad \text{Im}\left(\sum_{k=0}^{n-1} \left(e^{i\theta}\right)^k\right) = \sum_{k=0}^{n-1} \sin(k\theta)$$

Thus, if we make the substitution: $z = e^{i\theta}$, we find:

$$\begin{aligned}
\sum_{k=0}^{n-1} \left(e^{i\theta}\right)^k &= \frac{1-e^{i\theta n}}{1-e^{i\theta}} = \frac{1-z^n}{1-z} \\
&= \left(\frac{1-z^n}{1-z}\right)\left(\frac{\overline{1-z}}{\overline{1-z}}\right) \\
&= \left(\frac{1-z^n}{1-z}\right)\left(\frac{1-\bar{z}}{1-\bar{z}}\right) \\
&= \frac{1-z^n-\bar{z}+z^{n-1}(z\bar{z})}{1-(z+\bar{z})+z\bar{z}} \\
&= \frac{1-z^n-\bar{z}+z^{n-1}}{1-(z+\bar{z})+1} \\
&= \frac{1-z^n-(\bar{z})+z^{n-1}}{2-2\mathrm{Re}(z)}
\end{aligned}$$

And therefore,

$$\begin{aligned}
\sum_{k=0}^{n-1} \cos(k\theta) &= \mathrm{Re}\left(\sum_{k=0}^{n-1} \left(e^{i\theta}\right)^k\right) \\
&= \mathrm{Re}\left(\frac{1-z^n-(\bar{z})+z^{n-1}}{2-2\mathrm{Re}(z)}\right) \\
&= \mathrm{Re}\left(\frac{1-\cos(n\theta)-i\sin(n\theta)-\cos(\theta)-i\sin(\theta)+\cos((n-1)\theta)+i\sin((n-1)\theta)}{2-2\cos(\theta)}\right) \\
&= \frac{1-\cos(n\theta)-\cos(\theta)+\cos((n-1)\theta)}{2-2\cos(\theta)} \\
&= \frac{1}{2} - \frac{\cos(n\theta)-\cos((n-1)}{2-2\cos(\theta)}
\end{aligned}$$

And similarly for sin:

$$\begin{aligned}
\sum_{k=0}^{n-1} \sin(k\theta) &= \mathrm{Im}\left(\sum_{k=0}^{n-1} \left(e^{i\theta}\right)^k\right) \\
&= \mathrm{Im}\left(\frac{1-z^n-(\bar{z})+z^{n-1}}{2-2\mathrm{Re}(z)}\right) \\
&= \mathrm{Im}\left(\frac{1-\cos(n\theta)-i\sin(n\theta)-\cos(\theta)-i\sin(\theta)+\cos((n-1)\theta)+i\sin((n-1)\theta)}{2-2\cos(\theta)}\right) \\
&= -\frac{\sin(n\theta)+\sin(\theta)-\sin((n-1)\theta)}{2-2\cos(\theta)}
\end{aligned}$$

□

(b) *Proof.* We prove this fact directly:

$$\begin{aligned}
\frac{1}{n}\sum_{k=0}^{n-1} X(\omega_k)e^{-2\pi\omega_k s} &= \frac{1}{n}\sum_{k=0}^{n-1}\sum_{t=1}^{n} x_t e^{2\pi i \frac{k}{n} t} e^{-2\pi \frac{k}{n} s} \\
&= \frac{1}{n}\sum_{k=0}^{n-1}\left(x_t e^{2\pi i \frac{k}{n}(t-s)}\Big|_{t=s} + \sum_{t=1,t\neq s}^{n} x_t e^{2\pi i \frac{k}{n}(t-s)}\right) \\
&= \frac{1}{n}\sum_{k=0}^{n-1} x_s + \frac{1}{n}\sum_{k=0}^{n-1}\sum_{t=1,t\neq s}^{n} x_t e^{2\pi i \frac{k}{n}(t-s)} \\
&= \frac{1}{n}\sum_{k=0}^{n-1} x_s + \frac{1}{n}\sum_{t=1,t\neq s}^{n}\sum_{k=0}^{n-1} x_t e^{2\pi i \frac{k}{n}(t-s)} \\
&= \frac{1}{n}(nx_s) + \frac{1}{n}\sum_{t=1,t\neq s}^{n}\left(\frac{1-e^{2\pi i(t-s)}}{1-e^{2\pi i(t-s)n^{-1}}}\right) \\
&\stackrel{*}{=} x_s
\end{aligned}$$

Where the last step, $\stackrel{*}{=}$, is because $\forall t \in \{1,\dots,n\}$ $s \geq 1$ and $s \neq t$, $0 < 2\pi i(t-s)n^{-1} < 2\pi i$ and $2\pi i(t-s) \in 2\pi i m$ where $m \in \mathbb{Z}$. It is thus the case that for all elements within the domain of the final right-hand summation, they equate to $\frac{0}{z} = 0$ where $z \neq 0, z \in \mathbb{C}$, and hence why that summation evaluates to zero. □

1.3 The Spectral Density of a Stationary Stochastic Process

Suppose that $\{X_t\}$ is a stationary stochastic process with spectral density function $f_X(\omega)$ and define $\{Y_t\}$ by:

$$Y_t = \alpha\sum_{u=0}^{\infty}(1-\alpha)^u X_{t-u}, \quad 0 < \alpha < 1$$

a) Show that $\{Y_t\}$ can be defined recursively by:

$$Y_t = \alpha X_t + (1-\alpha)Y_{t-1}$$

(This is commonly referred to as **exponential smoothing**.)

b) Let $f_Y(\omega) = |\Gamma(\omega)|^2 f_X(\omega)$ be the spectral density function of $\{Y_t\}$. Evaluate $|\Gamma(\omega)|^2$ and determine for which frequencies $|\Gamma(\omega)|^2$ is largest and smallest. What is the effect of varying α?

(a) *Proof.* We prove this directly:

$$\begin{aligned}
Y_t &= \alpha \sum_{u=0}^{\infty} (1-\alpha)^u X_{t-u} \\
&= \alpha X_t + \alpha \sum_{u=1}^{\infty} (1-\alpha)^u X_{t-u} \\
&= \alpha X_t + \alpha \sum_{s=0}^{\infty} (1-\alpha)^{s+1} X_{t-s-1} \\
&= \alpha X_t + (1-\alpha) \left(\alpha \sum_{s=0}^{\infty} (1-\alpha)^s X_{t-s-1} \right) \\
&= \alpha X_t + (1-\alpha) Y_{t-1}
\end{aligned}$$

□

(b) **Answer:** From the filtering theorem, we have:

$$\begin{aligned}
\Gamma(\omega) &= \alpha \sum_{u=0}^{\infty} (1-\alpha)^u e^{2\pi i \omega u} \\
&= \frac{\alpha}{1-(1-\alpha)e^{2\pi i \omega}}
\end{aligned}$$

$$\begin{aligned}
\implies |\Gamma(\omega)|^2 &= \Gamma(\omega) \cdot \overline{\Gamma(\omega)} \\
&= \frac{\alpha(1-(1-\alpha)e^{-2\pi i \omega})}{(1-(1-\alpha)e^{2\pi i \omega})(1-(1-\alpha)e^{-2\pi i \omega})} \cdot \overline{\Gamma(\omega)} \\
&= \frac{\alpha\big(1-(1-\alpha)e^{-2\pi i \omega}\big)}{\big(1+(1-\alpha)^2 - 2\cos(2\pi i \omega)\big)} \cdot \frac{\alpha\big(1-(1-\alpha)e^{2\pi i \omega}\big)}{\big(1+(1-\alpha)^2 - 2\cos(2\pi i \omega)\big)} \\
&= \frac{\alpha^2\big(1+(1-\alpha)^2 - 2\cos(2\pi i \omega)\big)}{\big(1+(1-\alpha)^2 - 2\cos(2\pi i \omega)\big)^2} \\
&= \frac{\alpha^2}{1+(1-\alpha)^2 - 2\cos(2\pi i \omega)}
\end{aligned}$$

Now, we take the derivative of $|\Gamma(\omega)|^2$ with respect to ω to maximize / minimize $|\Gamma(\omega)|^2$:

$$\begin{aligned}
\frac{\partial |\Gamma(\omega)|^2}{\partial \omega} &= \frac{\partial}{\partial \omega} \left(\frac{\alpha^2}{1+(1-\alpha)^2 - 2\cos(2\pi i \omega)} \right) \\
&= \left(\frac{\alpha^2}{1+(1-\alpha)^2 - 2\cos(2\pi i \omega)} \right)^2 \cdot 2\sin(2\pi i \omega) \cdot 2\pi i
\end{aligned}$$

And therefore, we can see that $|\Gamma(\omega)|^2$ is maximized at $\omega = 0$, and minimized at $\omega = \frac{1}{2}$, since $|\Gamma(0)|^2 = 0$ and $\Gamma\big(\frac{1}{2}\big) = \frac{\alpha^2}{2-2\alpha+\alpha^2}$.

Therefore, we see that the amplitude of the frequency decreases as $\alpha \to 0$, and increases as $\alpha \to 1$.

1.4 Homogeneous Linear Recurrence Relations

Suppose that $\{X_t\}$ is an ARIMA(p,d,q) process and define $\hat{X}_{n+s}$ to be the best linear predictor of X_{n+s} given $X_n = x_n, X_{n-1} = x_{n-1}, \dots$, then:

$$\sigma^2(s) = \sigma^2 \sum_{u=0}^{s-1} \psi_u^2$$

where σ^2 is the white noise variance and $\{\psi_u\}$ are defined by the identity:

$$\sum_{u=0}^{\infty} \psi_u z^u = \frac{1 + \sum_{u=1}^{q} \beta_u z^u}{(1-z)^d (1 - \sum_{u=1}^{p} \phi_u z^u)}$$

(a) Suppose that $\{X_t\}$ is an ARIMA$(1,1,1)$ process with $\sigma^2 = 1$, $\phi_1 = 0.95$ and $\beta_1 = 0.2$. Evaluate $\sigma^2(s)$ for $s = 1, \dots, 10$. (You may want to write a simple program in R to do the computations.)

(b) Suppose that $\{X_t\}$ is an ARMA(p,q) process whose parameters satisfy the usual conditions. Show that $\sigma^2(s) \to \text{Var}(X_t)$ as $s \to \infty$. (Hint: Show that X_t can be written as

$$X_t = \mu + \sum_{u=0}^{\infty} \psi_u \varepsilon_{t-u}$$

where $\{\varepsilon_t\}$ is white noise).

(a) **Answer:** Let us begin by performing a little algebra on the identity which implicitly solves for ψ_i to make it such that it explicitly solves for ψ_i. However, to do this, we need a lemma:

Lemma. 1.1: Ahlfors: Complex Analysis Chpt 5: Thm 3 - Consequences

If $f(z)$ and $g(z)$ are analytic in the region Ω_1 and Ω_2, containing z_0, then both have the standard Taylor Series about z_0 ($f(z) = \sum_{n=0}^{\infty} a_n z^n$ and $g(z) = \sum_{n=0}^{\infty} b_n z^n$), and their product is determined by:

$$f(z)g(z) = \sum_{n=0}^{\infty} \left(\sum_{j=0}^{n} (a_j b_{n-j}) \right) z^n$$

And is analytic on $\Omega \geq \Omega_1 \cap \Omega_2$. Furthermore, the same derivation may be applied to $f(z)/g(z)$ provided $g(0) = b_0 \neq 0$.

Now, we return to the question at hand. We may now perform the following algebra on $\{\psi_u\}$ as this identity trivially satisfies our assumptions in the lemma.

$$
\begin{aligned}
\sum_{u=0}^{\infty} \psi_u z^u &= \frac{1 + \sum_{u=1}^{q} \beta_u z^u}{\big(1-z\big)^d \big(1 - \sum_{u=1}^{p} \phi_u z^u\big)} \\
\iff \left(\sum_{u=0}^{\infty} \psi_u z^u\right)\left(\big(1-z\big)^d \big(1 - \sum_{u=1}^{p} \phi_u z^u\big)\right) &= 1 + \sum_{u=1}^{q} \beta_u z^u
\end{aligned}
$$

Now in recalling our assumption of $p = d = q = 1$, we begin simplifying the left hand side above:

$$
\begin{aligned}
\left(\sum_{u=0}^{\infty} \psi_u z^u\right)\left(\big(1-z\big)^{(1)} \big(1 - \sum_{u=1}^{(1)} \phi_u z^u\big)\right) &= (1-z)(1-\phi_1 z)\left(\sum_{u=0}^{\infty} \psi_u z^u\right) \\
&= \big(1 - (1+\phi_1)z + \phi_1 z^2\big)\left(\sum_{u=0}^{\infty} \psi_u z^u\right) \\
&= \psi_0 + \psi_1 z + \sum_{u=2}^{\infty} \psi_u z^u - (1+\phi_1)\psi_0 z - \sum_{u=2}^{\infty}(1+\phi_1)\psi_{u-1} z^u + \sum_{u=2}^{\infty} \phi_1 \psi_{u-2} z^u \\
&= \psi_0 + \big(\psi_1 - (1+\phi_1)\psi_0\big)z + \sum_{u=2}^{\infty}\Big(\psi_u - \psi_{u-1} + \phi_1(\psi_{u-2} - \psi_{u-1})\Big)z^u
\end{aligned}
$$

And as such, we now look at the right hand side of our equation:

$$
1 + \sum_{u=1}^{(1)} \beta_u z^u = 1 + \beta_1 z
$$

And this therefore implies the following equations by equating the coefficients of our two equations:

$$
\begin{aligned}
&(1) \quad \psi_0 = 1 \\
&(2) \quad \psi_1 - (1+\phi_1)\psi_0 = \beta_1 \\
&(3) \quad \psi_u - \psi_{u-1} + \phi_1(\psi_{u-2} - \psi_{u-1}) = 0 \qquad\qquad \forall u \geq 2
\end{aligned}
$$

Solving for these linear equations, we find that:

$$
\psi_0 = 1, \quad \psi_1 = 1 + \beta_1 + \phi_1, \quad \psi_u = (1+\phi_1)\psi_{u-1} - \phi_1 \psi_{u-2}
$$

Instead of writing a code to solve for ψ_u, $u = 1, \ldots, 10$, we would like to attempt this algebraically. To do so, we need another lemma:

Lemma. 1.2: Solving homogeneous linear recurrence relations with constant coefficients (Yule-Walker Equations)

If $a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_d a_{n-d}$, then the roots of the polynomial:

$$P_d(t) = t^d - c_1 t^{t-1} - \cdots - c_d$$

which we shall call $r_1, \ldots, r_d$, will serve fundamental for the following explicate formula of a_n:

$$a_n = k_1 r_1^n + \cdots + k_d r_d^n$$

Where $k_1, \ldots, k_d$ are chosen to satisfy our initial conditions of $a_0, \ldots, a_{d-1}$.

We can now apply this to ψ_u, solving for the roots of the equation of $t^2 - (1+\phi_1)t + \phi_1 = 0$ are $r_1 = 1$, $r_2 = \phi_1$, and hence solving for the k_1, k_2 which yield our explicit formula, we have:

$$k_1 = \frac{1+\beta_1}{1-\phi}, \quad k_2 = \frac{\beta_1 + \phi_1}{\phi_1 - 1}$$

$$\implies \psi_n = \frac{1+\beta_1}{1-\phi_1} + \phi_1^n \left(\frac{\beta_1 + \phi_1}{\phi_1 - 1} \right), \quad \forall n \in \mathbb{N}$$

And as such, we have now found a_n $\forall n \in \mathbb{N}$. Thus, plugging in $\phi_1 = 0.95$, $\beta_1 = 0.2$, we have:

$$\psi_0 = 1, \quad \psi_1 = 1 + 0.95 + 0.2 = 2.15, \quad \psi_2 = \frac{1+0.2}{1-0.95} + (0.95)^2 \left(\frac{0.2+0.95}{0.95-1} \right) = 3.2425$$
$$\psi_3 = 4.280375, \quad \psi_4 = 5.26635625, \quad \psi_5 = 6.2030384375, \quad \psi_6 = 7.092886515625$$
$$\psi_7 = 7.93824218984375, \quad \psi_8 = 8.7413300803515625, \quad \psi_9 = 9.504263576333984375$$
$$\psi_{10} = 10.22905039751728515625$$

And since $\sigma = 1$, it follows immediately that:

$$\sigma^2(1) = 1, \quad \sigma^2(2) = 5.6225, \quad \sigma^2(3) = 16.1363, \quad \sigma^2(4) = 34.4579, \quad \sigma^2(5) = 62.1924$$
$$\sigma^2(6) = 100.67, \quad \sigma^2(7) = 150.979, \quad \sigma^2(8) = 213.995, \quad \sigma^2(9) = 290.406, \quad \sigma^2(10) = 380.737$$

(b) *Proof.* We first prove that an ARMA(p,q) model can be written as an MA(∞) process since, firstly, if X_t follow and ARMA(p,q) process, then $\phi(B)X_t = \theta(B)\varepsilon_t$, where $\phi(z) = 1 - \sum_{u=1}^p \phi_u z^u$ and $\theta(z) = 1 + \sum_{u=1}^q \beta_u z^u$. Thus, if $p < \infty$, then $\exists$ and analytic function $g(z) := \phi(z)^{-1} = 1/(1 - \sum_{u=1}^p)$ (Corollary of Lemma 1.1). Therefore, we have the following:

$$\phi(B)X_t = \theta(B)\varepsilon_t \iff X_t = g(B)\theta(B)\varepsilon_t, \text{ where } g(z)\theta(z) = \sum_{u=0}^{\infty} c_u z^u, \quad c_u \in \mathbb{C}$$

We recognize this as indeed an MA(∞) process. As such, we would like to evaluate the variance of this ARMA$(p,q) \equiv$ MA(∞) process, which we can do quite easily by noting the following:

$$\begin{aligned}
\text{Var}(X_t) &= \text{Var}\left(\sum_{u=0}^{\infty} c_u \varepsilon_{t-u}\right) \\
&= \sigma^2 \sum_{u=0}^{\infty} c_u^2 + \sum_{u \neq v} c_u c_v \text{Cov}(\varepsilon_{t-u}, \varepsilon_{t-v}) \\
&= \sigma^2 \sum_{u=0}^{\infty} c_u^2 && \text{since Cov}(\varepsilon_{t-u}, \varepsilon_{t-v}) = 0 \ \forall u \neq v
\end{aligned}$$

Now, we need only recognize that in the question, $\psi_u := c_u$, and hence:

$$\implies \lim_{s\to\infty} \sigma^2(s) = \lim_{s\to\infty} \left(\sigma^2 \sum_{u=0}^{s-1} \psi_u^2\right) = \sigma^2 \sum_{u=0}^{\infty} \psi_u^2 = \text{Var}(X_t)$$

□

2 Applications

For a full copy of the evolution of the code used for the questions on applications, please visit my GitHub at the following link: https://github.com/jmostovoy/Time_Series. A copy of the final code is provided in the Appendix.

2.1 Correlograms and Periodogram

Daily Canadian/U.S. dollar exchange rates ($US/$CAN) from Jan. 2, 1997 to Dec. 29, 2000 are given in the file `dollar.txt` *on Blackboard. Analyze the data on the log-scale:*

```
> dollar <- scan("dollar.txt")
> dollar <- ts(log(dollar))
```

Define the first differences of the time series as follows:

```
> returns <- diff(dollar)
```

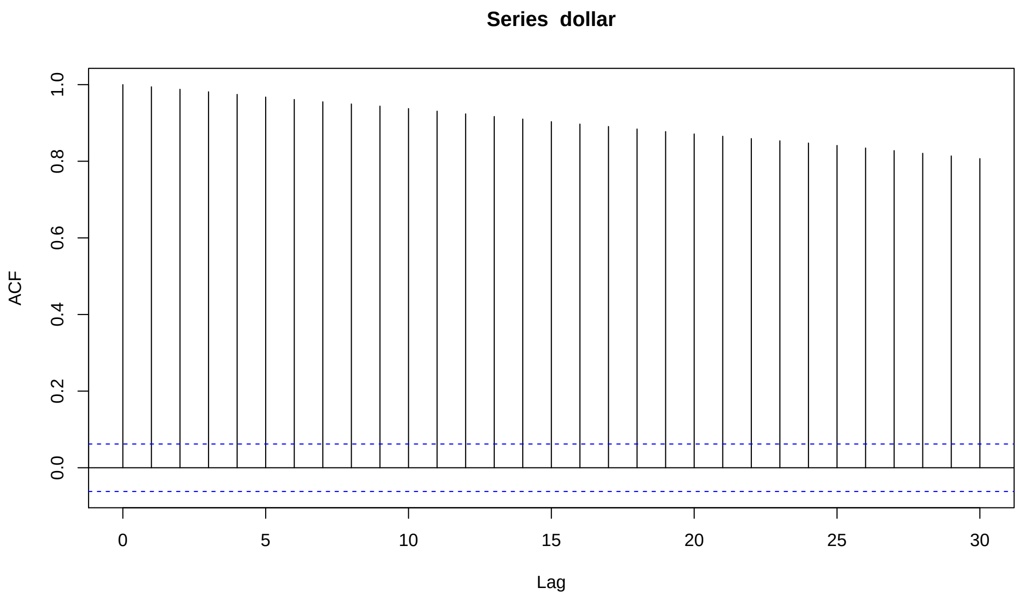
(a) *Plot the correlogram and periodogram of the orginal data (i.e.* `dollar`*).*

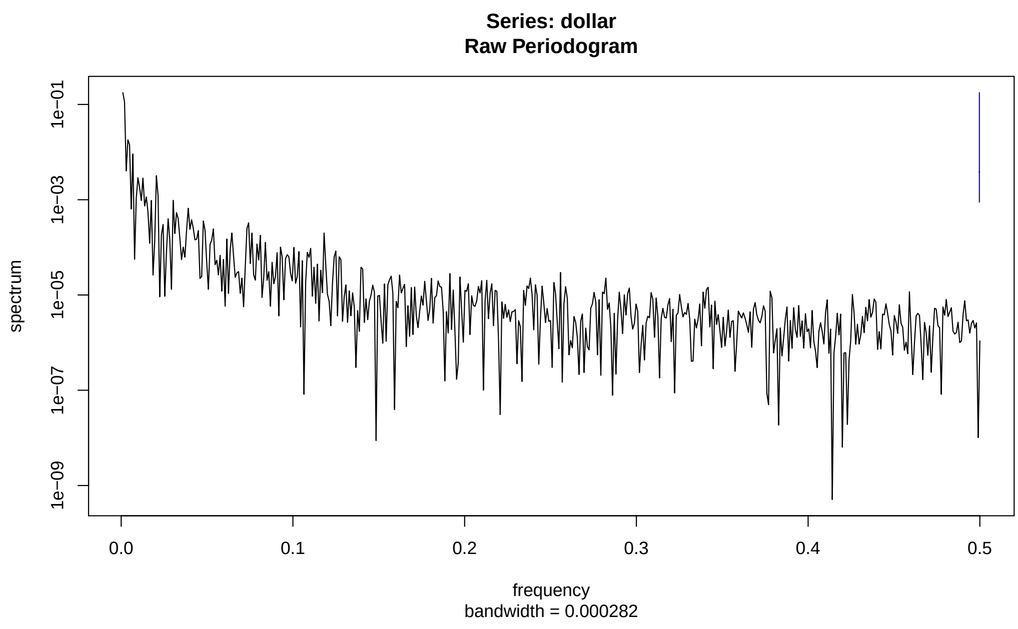
(b) *Plot the correlogram and periodogram of the first differences*

(c) *Comment on the results obtained in parts (a) and (b). In particular, how are the correlograms and periodograms different? A simple model for the logarithm of exchange rates is a random walk – are the correlograms and periodograms in (a) and (b) consistent with this model?*

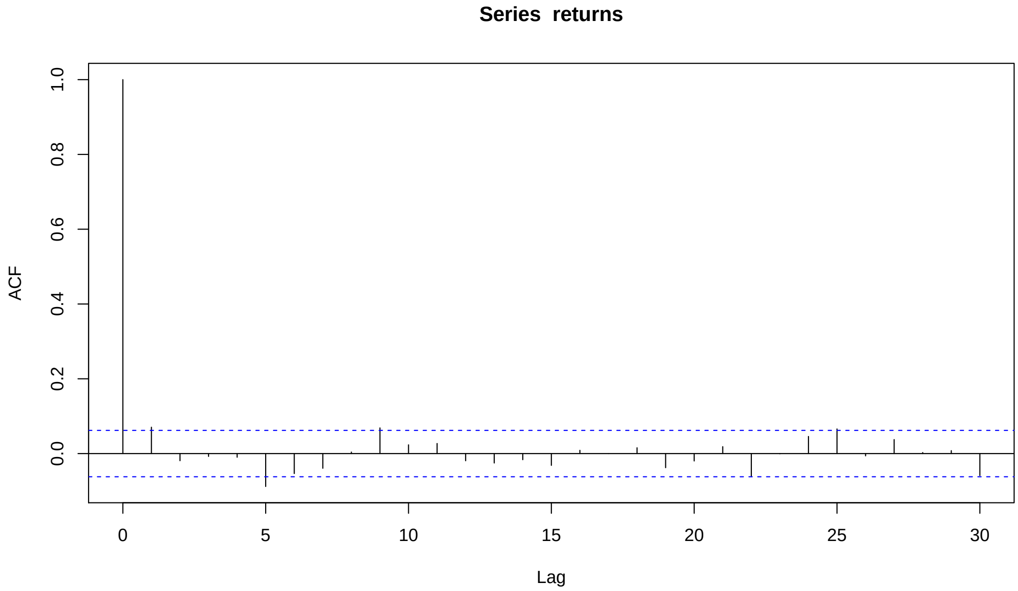
(d) *Now look at the correlogram and periodogram of the absolute values of the first differences (i.e.* `abs(returns)`*). Comment on the differences between the results for* `returns` *and* `abs(returns)`*, in particular, with respect to the applicability of the random walk model.*

(a) *Answer:*

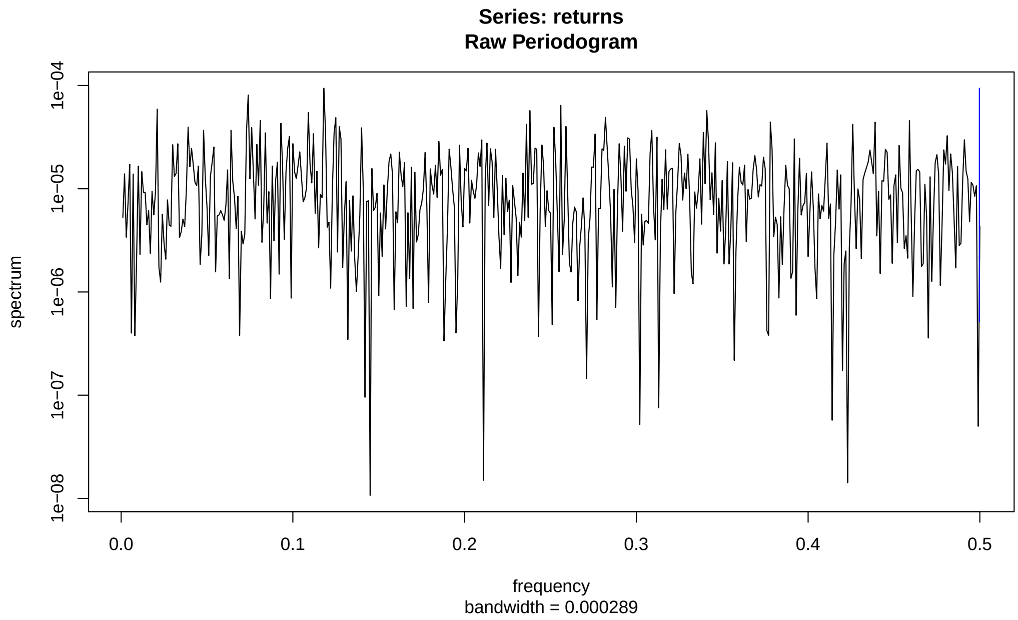

Series dollar
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag


Series: dollar
Raw Periodogram
spectrum
1e−01
1e−03
1e−05
1e−07
1e−09
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.000282

(b) *Answer:*



Series returns
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag



Series: returns
Raw Periodogram
spectrum
1e−04
1e−05
1e−06
1e−07
1e−08
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.000289

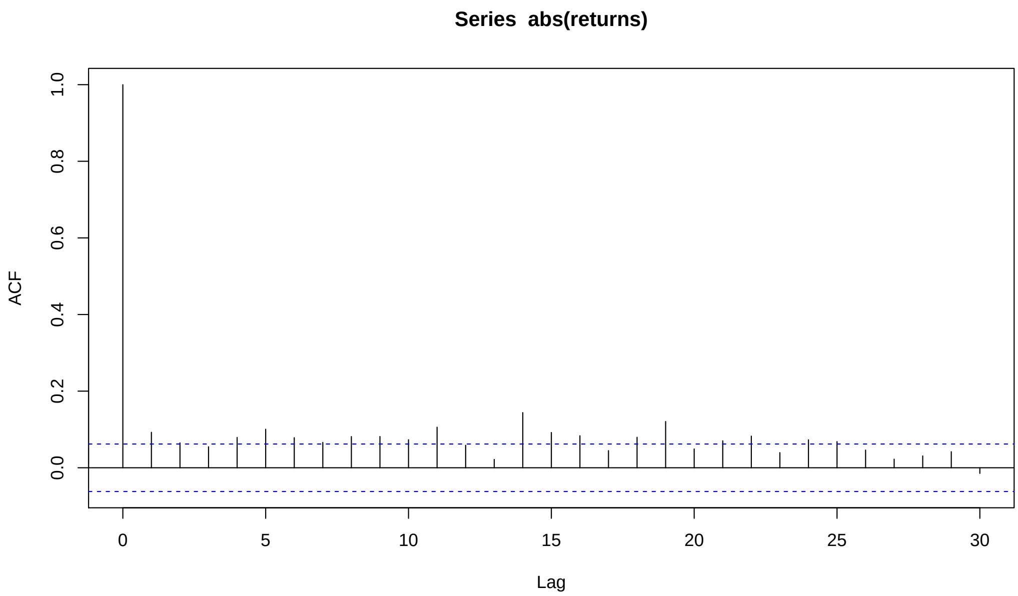
(c) *Answer:*

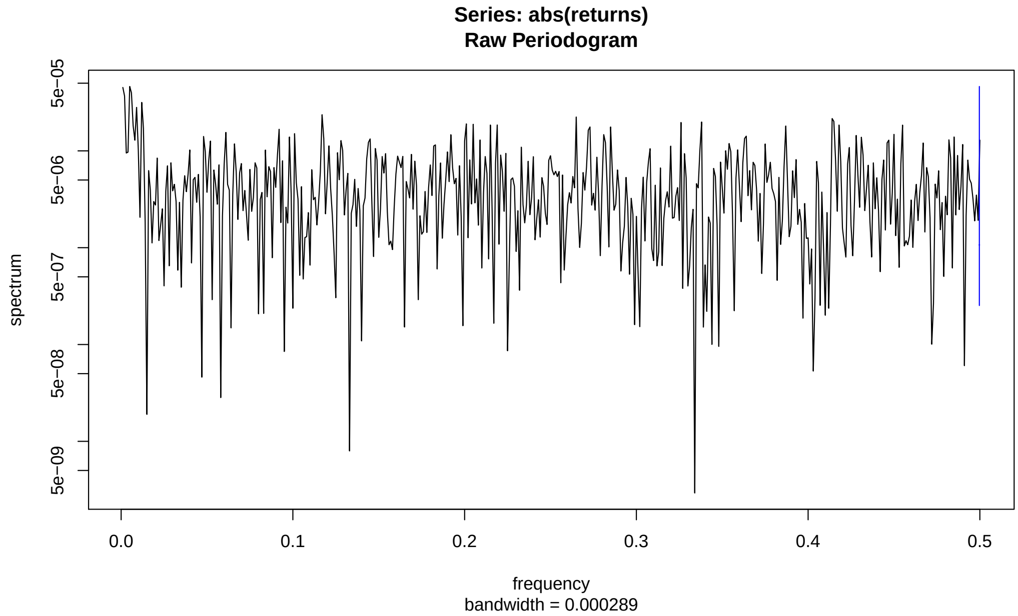
Naturally, the most immediate findings are that with respect to `dollar`, the correlogram decays very slowly, whereas for `returns`, the correlogram decays to 0 almost immediately.

For `dollar`'s Periodogram, it has a maximum near 0, and is generally decreasing as frequency increases. For `returns`'s Periodogram, it has no obvious maximum or minimum values (looks random around its mean).

As such, based on (a) and (b)'s periodogram and correlogram, we see that all the charecterstics described above do appear to be consistent with the random walk model. This is because the first difference of the random walk process is precisely white noise, which has no auto-correlation with past histories. Since we see this, and a constant periodigram for the first difference, we can make the claim that these correlograms and correlograms are indeed consistent.

(d) *Answer:*





Much like the correlogram and correlogram for `returns`, we see that `abs(returns)`'s correlogram decays to zero very quickly, while admitting slightly more significant possible auto-correlations for lags between 1 and 22. As for `abs(returns)`'s periodigram, there the only noticible difference in its shape is now having a small peak near the 0 frequency (whereas `returns`'s periodigram did not have this property). Another noticeable difference is that the mean spectrum on the `abs(returns)`'s periodigram is 5.2982×10^{-6}, in comparison to 1.1932×10^{-5} (I.e., avg_{spec}(`abs(returns)`)$\approx 0.445 \cdot$ avg_{spec}(`returns`). As such, because of the changes in the correlogram's and periodigram's attributes as just discussed, and since the first difference is supposed to behave a completely random behaviour, we see that the applicability of the random walk model is a little less now because we see a little too strong of serial correlation between the magnitude's of returns for lags / frequencies close to zero.

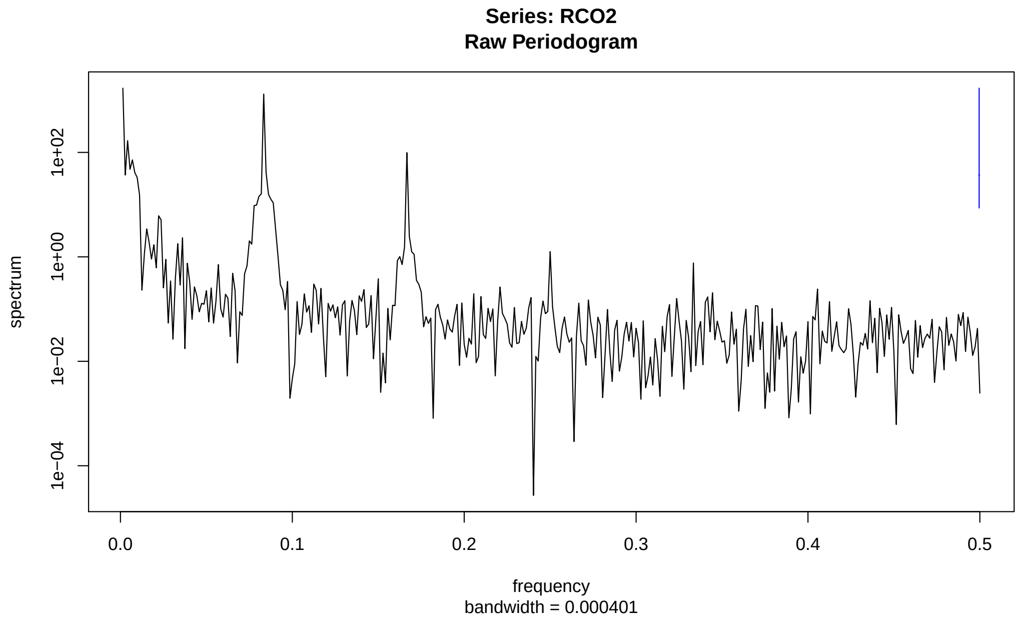
2.2 More Periodograms

Average monthly concentrations of carbon dioxide (CO_2) from March 1958 to December 2016 at Mauna Loa volcano in Hawaii are given in the file `CO2.txt`.

(a) *Plot the periodogram of the time series. At what frequencies are there peaks? To which features of the time series do these peaks correspond?*

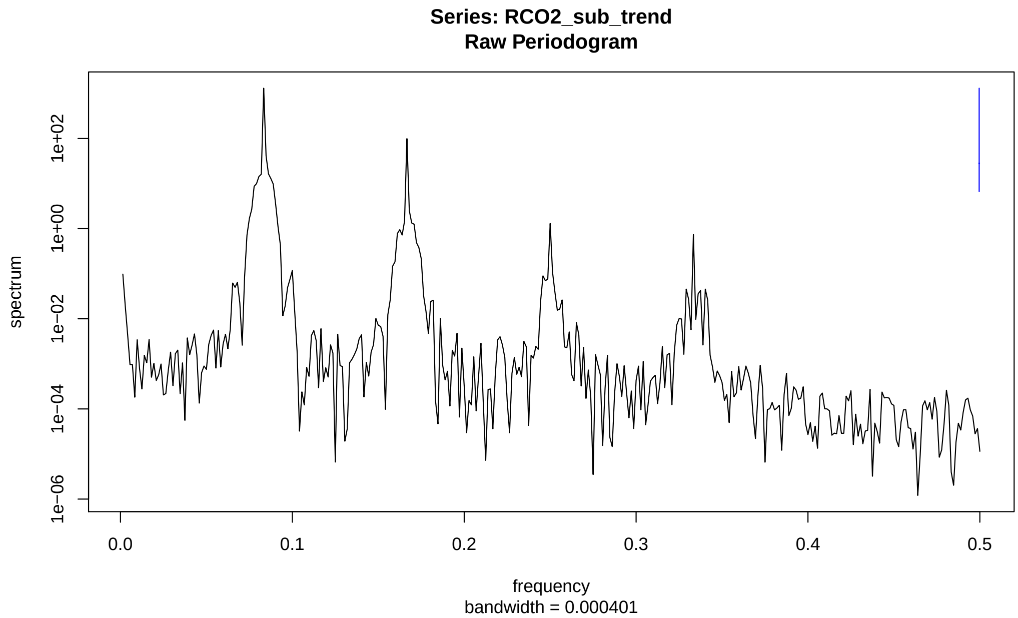
(b) *An estimate of the trend is given in the file* `CO2-trend.txt`. *Subtract the trend from the original data and look at the periodogram of the detrended data. Comment on the differences between the periodograms in parts (a) and (b). (It is useful here to overlay the two periodograms on the same plot.) In particular, how effective is the detrending in emphasizing the seasonality in the data?*

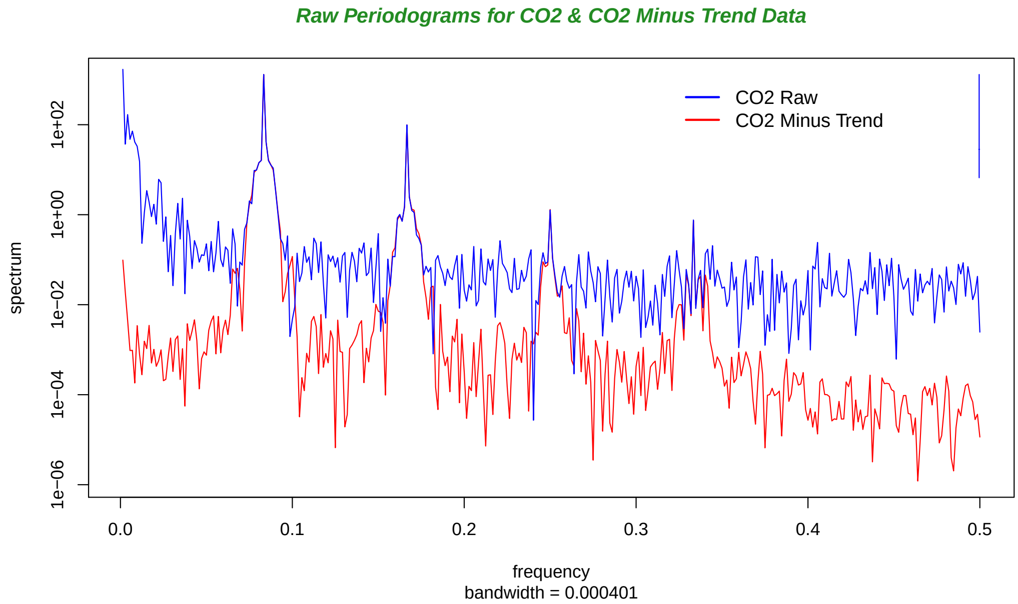
(a) *Answer:*



We thus see peaks at frequencies $f = 0, 0.0833, 0.1667, 0.2500, 0.3333$, which correspond to months $m = 0, 12, 24, 36, 48$. (Hence `CO2` is a yearly cyclical process, which has not been differenced yet, hence why $m = 0$ is included).

(b) *Answer:*





Naturally, here, one sees that the emphasis on the yearly cyclicality is much more pronounced on the `CO2 Minus Trend` Periodogram. The mean of the `CO2 Minus Trend` Periodogram is also lower than the `CO2 Raw` Periodogram, hence why the seasonal cyclicality is more pronounced. Also, since subtracting the trend is almost like a first differencing, we now have that `CO2 Minus Trend` Periodogram no longer has a pronounced peek near the 0 frequency.

2.3 Using the Periodogram, ACF (Corelogram) and PACF (Partial Corelogram) to Choose ARIMA Model

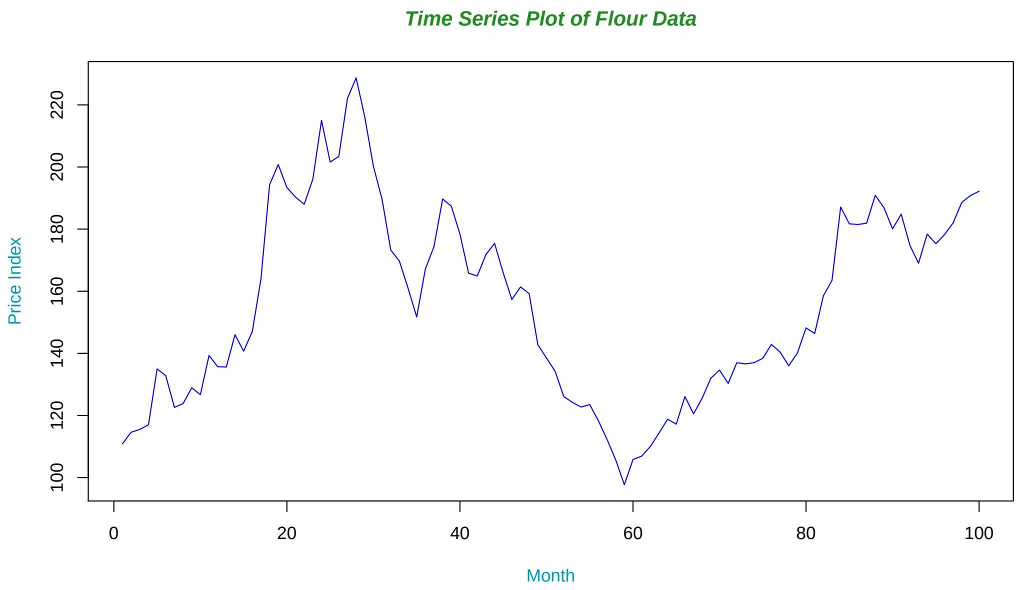
Monthly values of the Kansas City flour price index from August, 1972 to November, 1980 are contained in the file `flour.txt`.

(a) Look at a time series plot of the data. Are there any obvious trends and/or periodicities in the data? Would it be worthwhile to transform the data?

(b) Plot the correlogram, partial correlogram and periodogram of the data.

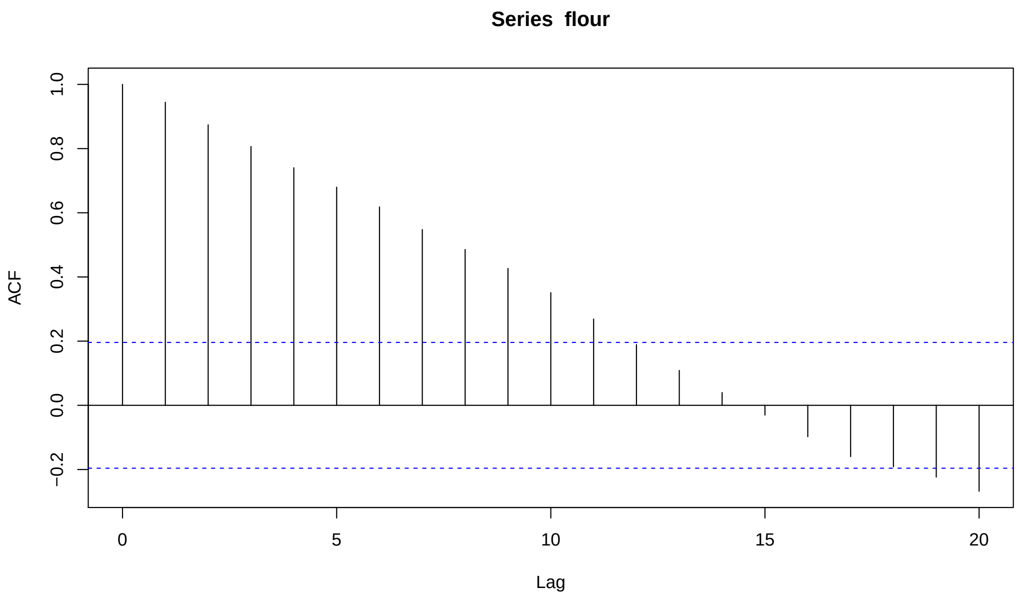
(c) Which of the following models seems to be most appropriate for these data: an autore- gressive model, a moving average model or a random walk (ARIMA$(0, 1, 0)$) model? Justify your answer.

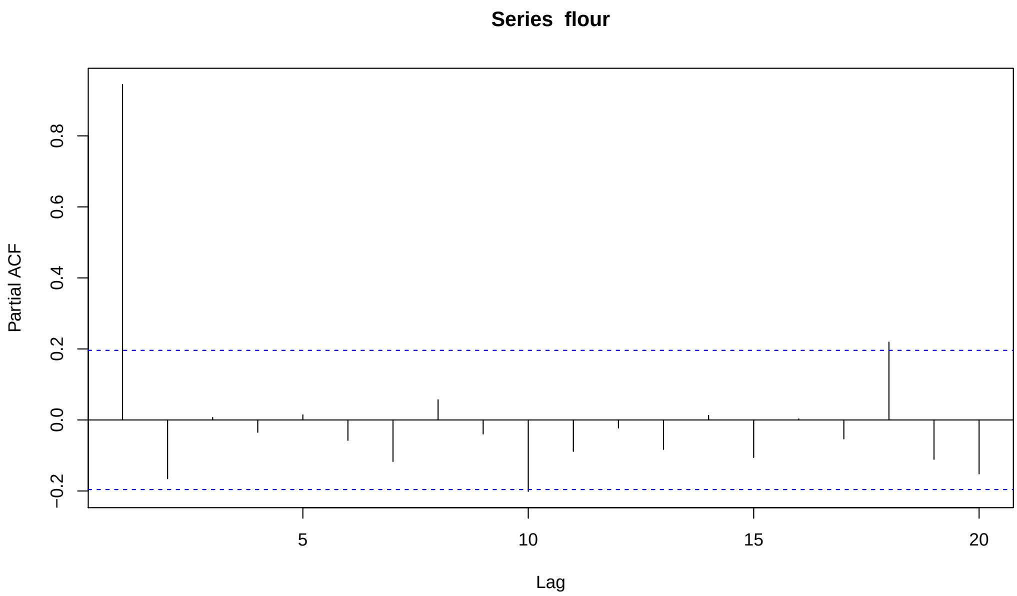
(a) **Answer:**

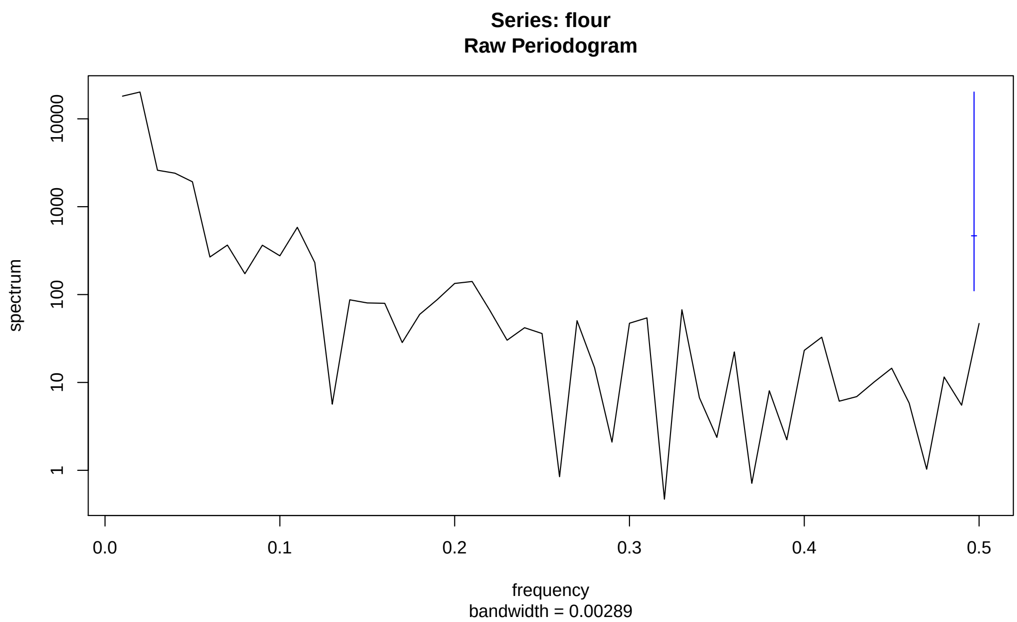


As we can see, there does not appear to be any hidden trends in which we could extract by transformations.

(b) **Answer:**







The R code used here is as follows:

```
    ####Preamble####

library(stats)

####Question 2####
```

```
#Part a
setwd("~/Documents/Time_Series")
flour <- scan(file="flour.txt")
flour <- ts(flour)
View(flour)

plot(flour,  type="l", col=4,lty=1, ann=FALSE)
title(xlab="Month", col.lab=rgb(0,0.6,.7))
title(ylab="Price Index" , col.lab=rgb(0,0.6,.7))
title(main="Time Series Plot of Flour Data",
      col.main="forestgreen", font.main=4)

#Part b

acf(flour)
spec.pgram(flour)
pacf(flour)
```

(c) **Answer:**
We use the following code in R to test for the best ARIMA model:

```
#Test for best Arima:
forecast::auto.arima(flour)
fitAR11<-arima(flour, order = c(1,0,1))
print(fitAR11)
Box.test(residuals(fitAR11), lag = 2, type = "Ljung-Box")
ar(flour)
```

And as such, we see that ARIMA(2,0,0) is the best possible fit for this data (i.e., AR(2)).

Also, the respective output for this code is:

```
> Box.test(flour)

Box-Pierce test

data:  flour
X-squared = 89.197, df = 1, p-value < 2.2e-16

> forecast::auto.arima(flour)
Series: flour
ARIMA(2,0,0) with non-zero mean

Coefficients:
         ar1      ar2  intercept
      1.2010  -0.2447   154.4943
s.e.  0.0963   0.0975    16.5668

sigma^2 estimated as 71.48:  log likelihood=-355.24
```

```
AIC=718.49   AICc=718.91   BIC=728.91
> fitAR11<-arima(flour, order = c(1,0,1))
> print(fitAR11)

Call:
arima(x = flour, order = c(1, 0, 1))

Coefficients:
         ar1     ma1  intercept
      0.9482  0.2665   154.4844
s.e.  0.0320  0.0993    17.3820

sigma^2 estimated as 69.1:  log likelihood = -355.08,  aic = 718.16
> Box.test(residuals(fitAR11), lag = 2, type = "Ljung-Box")

Box-Ljung test

data:  residuals(fitAR11)
X-squared = 0.01608, df = 2, p-value = 0.992

> ar(flour)

Call:
ar(x = flour)

Coefficients:
      1        2
 1.1007  -0.1655

Order selected 2  sigma^2 estimated as  104.8
```

2.4 ADF and White Noise Tests

Consider the \$US/\$CAN exchange rate data previously analyzed in Assignment #1; these data are in the file `dollar.txt` on Blackboard. As before, analyze the data on the log-scale. This time series appears to be non-stationary but has stationary first differences.

(a) Carry out the Augmented Dickey-Fuller (ADF) test of the null hypothesis that the data have a unit root for various lags. (To use the R function `adf.test`, you must load the R package `tseries`; `adf.test` can then be used as follows:

```
> library(tseries)
> adf.test(x,k=5)
```

where `x` is a vector or time series.) What is your conclusion?

(b) Do the first differences seem to be close to white noise? Use Bartlett's test and the Box-Pierce (Portmanteau) test to test this.

Note: A special R function has been written to carry out Bartlett's test: in R, type:

```
> source(\bartlett.txt")
> s <- bartlett(x)
```

where x is name of the vector or time series. The variable s will have two components: `s$stat` contains Bartlett's statistic and `s$pvalue` contains the p-value. The R function `Box.test` can be used for the Box-Pierce test; for example, `r <- Box.test(x,lag=10,type=\Ljung")` computes the statistic defined in class and its p-value (based on a χ^2 approximation).

(a) **Answer:** We do not have enough evidence to conclude that this process is free of unit roots. The p-values are reported below:

P-Values for ADF Test on Given Lags					
lags 1-5	0.504	0.514	0.494	0.493	0.561
lags 6-10	0.582	0.610	0.605	0.524	0.506
lags 11-15	0.489	0.502	0.527	0.513	0.549
lags 16-20	0.550	0.532	0.514	0.566	0.592

(b) **Answer:** Our Barlett Test suggests that this first difference is unlikely to be white noise due since at our found p-value (0.0408), we have strong evidence to conclude that the model is heteroskedastic, and hence unlikely to be a white noise process. We can see the same with the Box Tests, which all have low p-values as reported below, and hence the data is likely not independent of each other, and again refuting our claim that this is a white noise process:

P-Values for Box Test on Given Lags					
lags 1-5	0.026	0.069	0.144	0.239	0.021
lags 6-10	0.013	0.013	0.023	0.007	0.010
lags 11-15	0.014	0.019	0.024	0.034	0.037

2.5 Stationarity and (Further) Fitting an ARIMA Model

Monthly yields on short-term British government securities for a 21 year period are contained in a file `yield.txt` on Blackboard.

(a) Do the data appear to be non-stationary? Does a random walk (ARIMA$(0, 1, 0)$) model appear to fit the data? (Use formal and informal white noise tests here.)

(b) Using the arima function in R, fit ARIMA$(p, 1, p)$ models for $1 \leq p \leq 3$. Use AIC to determine the best fitting ARIMA$(p, 1, p)$ model. (If `z <- arima(....)` then `z$aic` contains the value of AIC for the model fitted.)

(c) Try a few other models and look at AIC for each of them. Which ARIMA$(p, 1, q)$ model seems to be best?

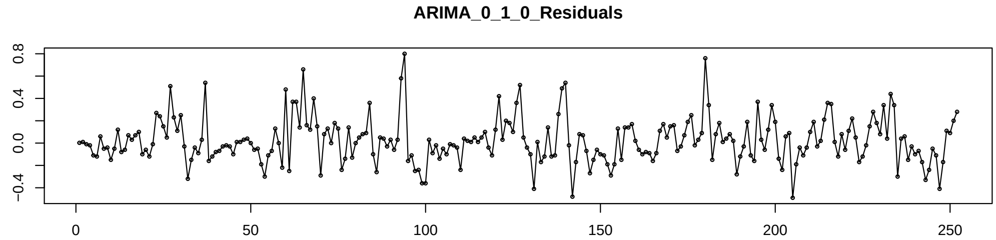
(d) For your best model, are the residuals close to a white noise process? Also check the normality of the residuals. (The function `tsdiag` is very useful here.)

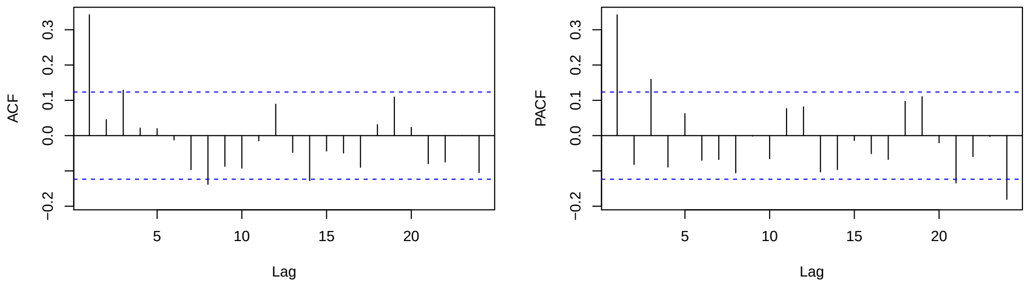
(a) **Answer:** To determine if the data is non-stationary, we use the ADF and Box Tests as formal tests, and look at the ACF and PACF as informal tests. The results from the ADF tests are:

P-Values for ADF Test on Given Lags					
lags 1-5	0.060	0.096	0.019	0.039	0.016
lags 6-10	0.026	0.043	0.090	0.083	0.130
lags 11-15	0.059	0.020	0.046	0.091	0.089
lags 16-20	0.122	0.202	0.067	0.014	0.010

And hence we would say that this test would be inconclusive. For the Box Test, for all lags, we have an infinitesimal (≈ 0) p-value. Therefore, it is most likely that the data is not independently distributed; they exhibit serial correlation, and hence non-stationarity.

For the first difference model (i.e., ARIMA$(0,1,0)$) and we report the findings:





P-Values for Box Test on Given Lags Times Scaled by 10^6					
lags 1-5	0.045	0.247	0.143	0.512	1.578
lags 6-10	4.434	4.000	1.159	1.214	1.119
lags 11-15	2.481	2.293	3.813	1.418	2.371

And our p-value for Bartlett Test is 7.931538^{-6}.

Therefore, we have that the ACF and PACF have significant lags, and both tests imply that this the ARIMA$(0,1,0)$ does not fit properly since its not stationary / not white noise.

(b) **Answer:** We see that the best ARIMA$(p,1,p)$, $1 \leq p \leq 3$ is ARIMA$(3,1,3)$ if we just consider the AIC Values ((3,1,3) has the lowest AIC). Explicitly, AIC values for ARIMA$(1,1,1)$, $(2,1,2)$, and $(3,1,3)$ are -119.3551, -119.6975, and -131.1786 respectively.

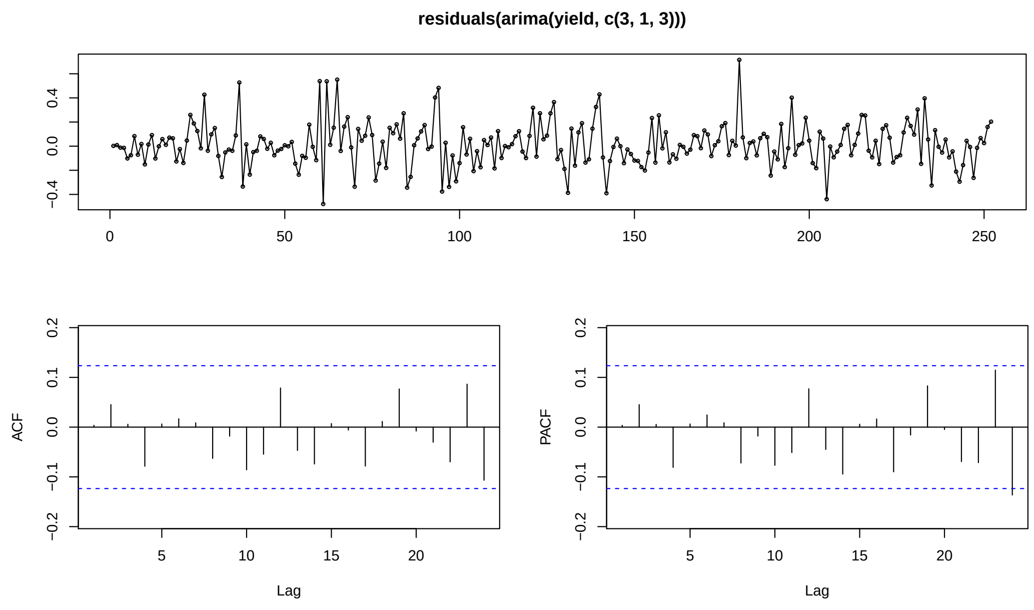
(c) **Answer:** We show the data:

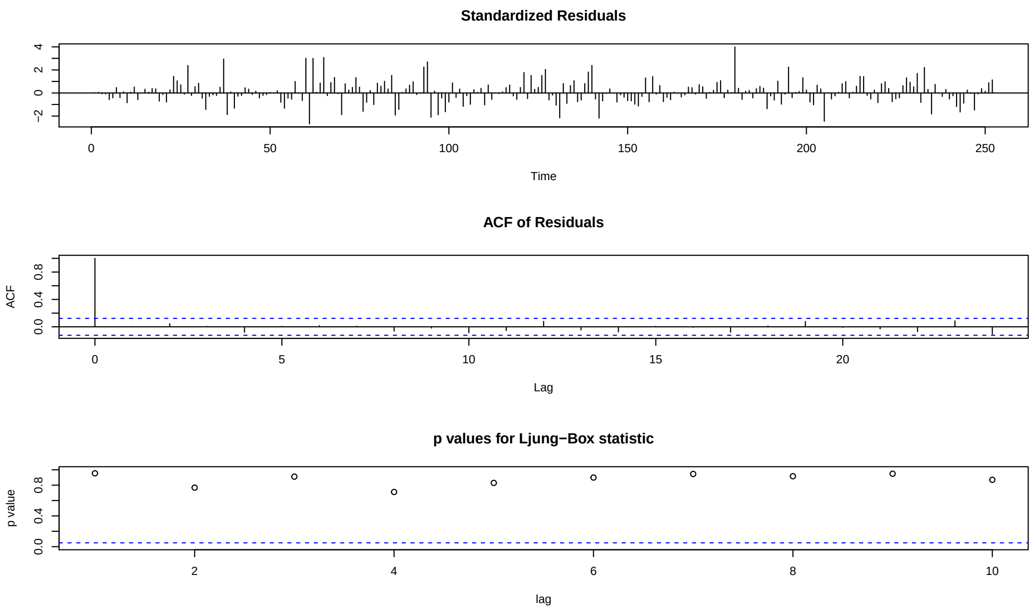
AIC for Different ARIMA$(p, 1, q)$ Models						
$p = P \setminus q = Q$	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
$p = 0$	-82.949	-119.520	-118.380	-121.650	-119.701	-118.225
$p = 1$	-113.913	-119.355	-118.922	-119.856	-117.950	-116.456
$p = 2$	-113.315	-120.148	-119.697	-118.079	-116.133	-131.848
$p = 3$	-118.236	-119.878	-118.009	-131.179	-133.764	-129.999
$p = 4$	-117.966	-118.033	-117.003	-131.256	-114.043	-136.778
$p = 5$	-117.051	-116.033	-114.035	-115.308	-127.478	6.000

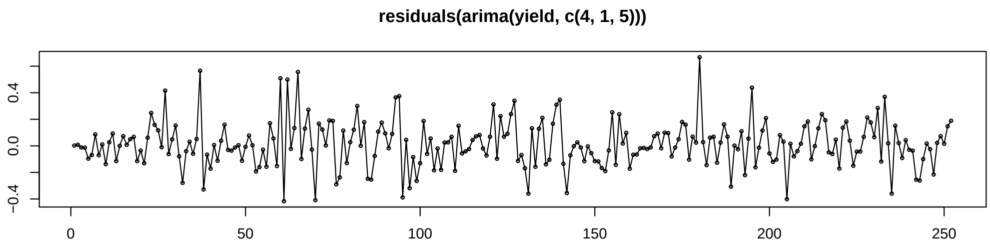
Thus the minimum AICs occur for ARIMA$(4, 1, 5)$ then ARIMA$(3, 1, 4)$ then ARIMA$(4, 1, 3)$ then ARIMA$(3, 1, 3)$. As such, If $p, q \in \{0, \ldots, 5\}$, then we would select ARIMA$(4, 1, 5)$, if $p, q \in \{0, \ldots, 4\}$, then the best model is ARIMA$(3, 1, 4)$ and if $p, q \in \{0, \ldots, 3\}$ then the best model is ARIMA$(3, 1, 3)$. However, since AIC tends to favour models with higher orders, it's probably best we stick to ARIMA$(3, 1, 3)$ for simplicity.

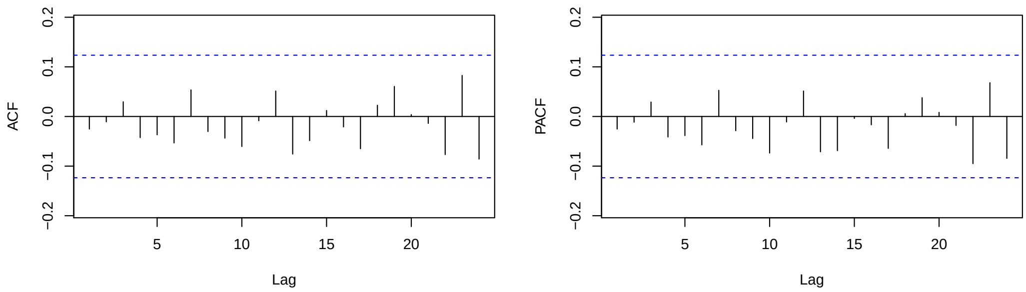
(d) **Answer:** We perform the necessary tests for both ARIMA$(4, 1, 5)$ and ARIMA$(3, 1, 3)$ for the reasons states in 2c. Furthermore, all our tests say that the residuals for both the ARIMA$(4, 1, 5)$ and ARIMA$(3, 1, 3)$ are very close to a white noise process for their p-values are all significant. Moreover, by looking at a QQ-Plot, we see that the tails of the distribution are too thick, and hence the residuals are most likely not normal. We report the supporting data for these claims:

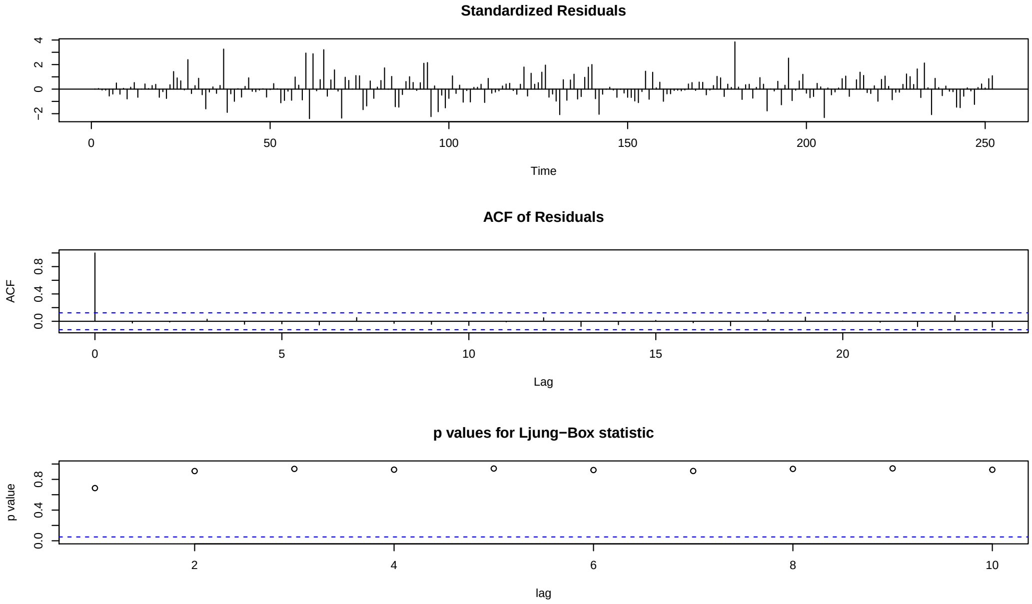
`bartlett_p_value(ARIMA(3,1,3)$residuals)`$=0.9269218$
`bartlett_p_value(ARIMA(4,1,5)$residuals)`$=0.957829$




Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung−Box statistic
p value
lag


residuals(arima(yield, c(4, 1, 5)))

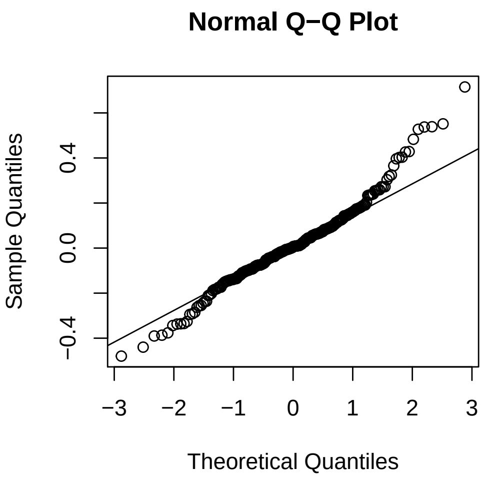

ACF
Lag
PACF
Lag

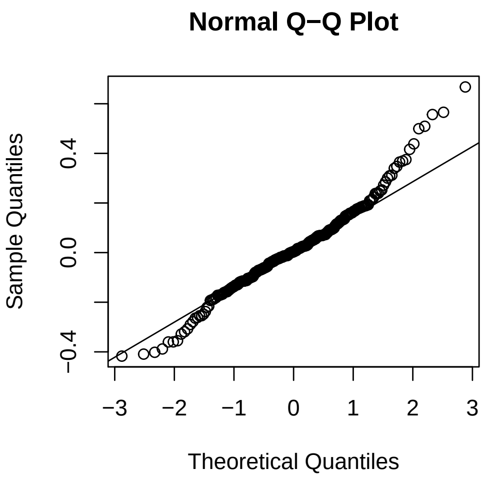


P-Values for Box Test on Given Lags for ARIMA(3,1,3)					
lags 1-5	0.954	0.768	0.911	0.712	0.829
lags 6-10	0.899	0.946	0.916	0.948	0.870
lags 11-15	0.867	0.805	0.821	0.776	0.831

P-Values for Box Test on Given Lags for ARIMA(4,1,5)					
lags 1-5	0.686	0.908	0.936	0.926	0.941
lags 6-10	0.922	0.909	0.936	0.943	0.926
lags 11-15	0.955	0.953	0.917	0.921	0.947

QQ Plot for `ARIMA(3,1,3)$residuals` QQ Plot for `ARIMA(4,1,5)$residuals`





2.6 Fitting a SARIMA Model

Data on the number of monthly traffic fatalities in Ontario from 1960 to 1974 are contained in a file `fatalities.txt` on Blackboard. You may want to analyze the logs of the data.

a) Fit a seasonal ARIMA (SARIMA) model to the data using the following steps:

   i) Difference the data; that is, let $y_t = \nabla\nabla_{12}x_t$ where x_t is the original data.

   ii) Plot the correlogram and partial correlogram of y_t; choose the AR and MA orders in the SARIMA model based on these plots.

   iii) Use arima to estimate the parameters in your model.

   iv) Check that the residuals from your model are close to white noise.

   (You should find that a SARIMA$(0,1,1) \times (0,1,1)_{12}$ model fits well although you may want to try other models.)

b) Use the R function predict to forecast the number of traffic fatalities for each month of 1975. Give 95% prediction limits for your forecasts. (If r is the output from the "best" SARIMA model, `r1 <- predict(r,n.ahead=12)` will give forecasts and prediction standard errors for each month — `r1$pred` contains the forecasts while `r1$se` contains the standard errors.)

c) Use the data from 1960 to 1973 to estimate the parameters of your "best" model from part (a). Then use these parameter estimates to forecast the number of fatalities for 1974 and obtain 95% prediction limits. How do the true number of fatalities compare to these forecasts?

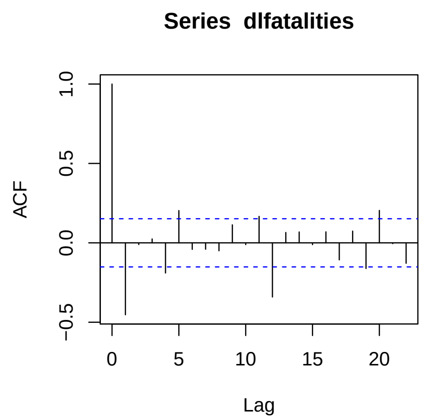
Note: Questions 4a-c are all analyzed on a log scale as recommended.

a) **Answer:**

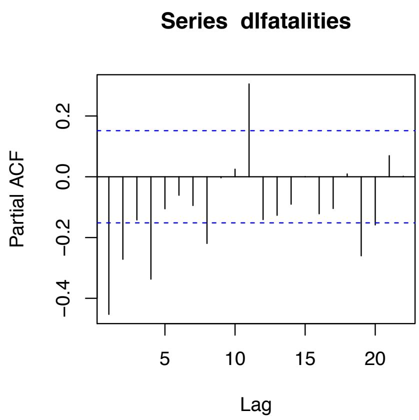
   i) Please reference our code for how we carried out the differencing.

   ii) We first recall the following: An MA(q) process will have a correlogram that cuts off after q lags, and a partial correlogram that tails off. Thus, we show both periodigram and correlogram:

ACF Plot for y_t



PACF Plot for y_t



Both the ACF and the PACF show very significant peaks at lag 1 and 12. Furthermore, the PACF displays autocorrelation for many lags. As such, and recalling how an MA v. AR process' ACF and PACF can indicate parameters, we feel that a seasonal moving average and an ordinary MA(1). Thus, we expect the SARIMA model to have the model SARIMA$(0,1,1) \times (0,1,1)_{12}$.

iii) The `Arima` function also predicts SARIMA$(0,1,1) \times (0,1,1)_{12}$ as this model has the lowest AIC value:

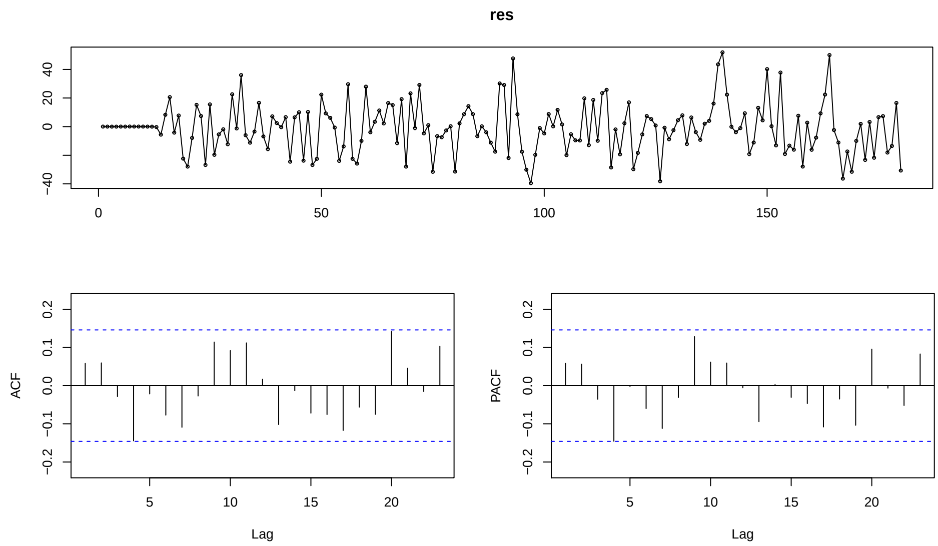
AIC for Different SARIMA$(p,1,q) \times (P,1,Q)_{12}$ Models						
$p = P \setminus q = Q$	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
$p = 0$	1588.528	1460.227	1462.446	1463.302	1461.993	1465.509
$p = 1$	1520.892	1462.108	1463.756	1464.826	1462.575	1466.192
$p = 2$	1500.515	1462.415	1466.444	1459.069	1461.975	1456.251
$p = 3$	1483.438	NA	NA	NA	NA	NA
$p = 4$	1472.859	1465.269	1468.807	NA	NA	NA

We can also double check these results with the `auto.arima` function, and find as well that `auto.arima`(y_t)=(0,0,1) $\implies$ SARIMA$(0,1,1) \times (0,1,1)_{12}$ fits our data best.

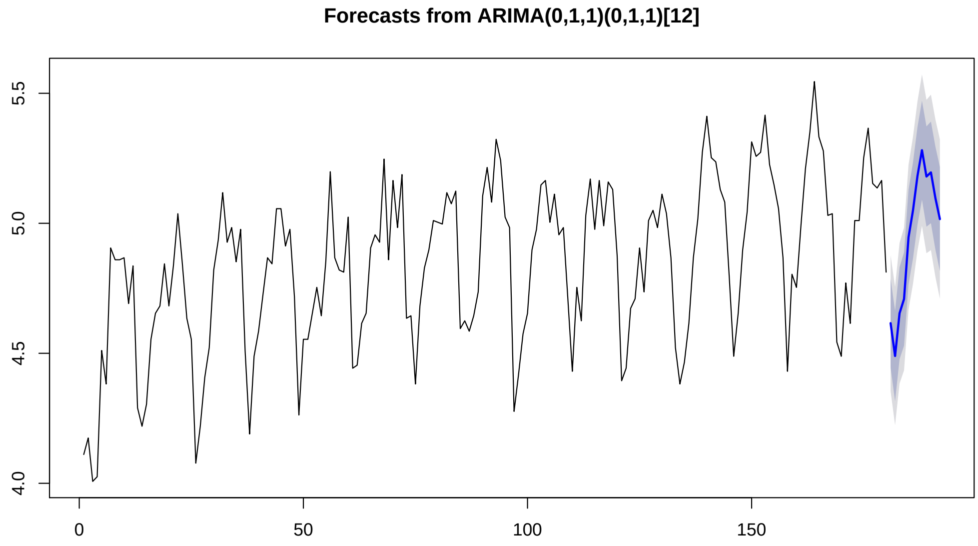
iv) Our formal tests for residuals being close to white noise have low p-values and our informal tests for white noise have insignificant values, hence the residuals from our model are indeed close to white noise. Our data for which we base this on is shown below:

P-Values for Box Test on Different Lags for Residuals of SARIMA$(0,1,1) \times (0,1,1)_{12}$					
lags 1-5	0.978	0.666	0.812	0.153	0.236
lags 6-10	0.290	0.257	0.348	0.287	0.310
lags 11-15	0.367	0.403	0.222	0.256	0.232

`bartlett_p_value(SARIMA`$(0,1,1) \times (0,1,1)_{12}$`$residuals)`=0.7054855



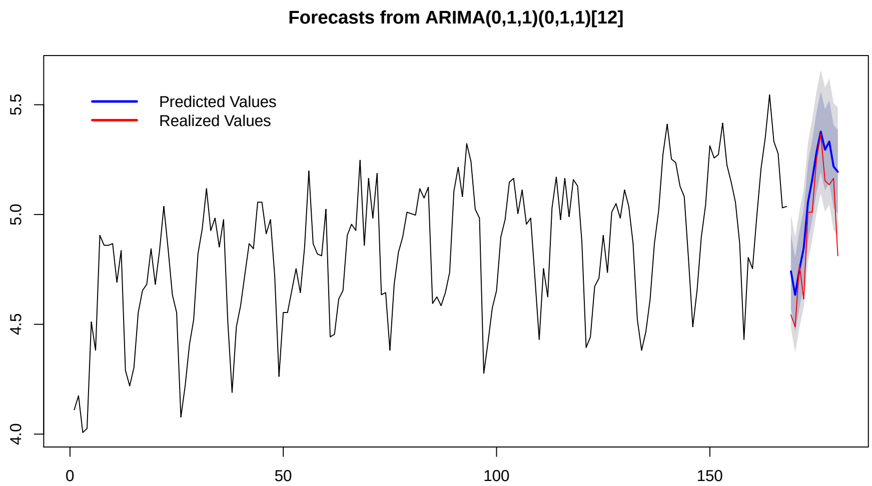
b) **Answer:** We show the following graph as our final result (the dark blue are predicted values, the darker shade of blue is the 80% prediction limits, and the lighter blue are the 95% prediction limits):



We also summarize the required data in the following table:

Prediction Values and Limits for our SARIMA$(0,1,1) \times (0,1,1)_{12}$ Model												
low 95%	4.35	4.22	4.38	4.43	4.67	4.77	4.90	4.99	4.89	4.90	4.79	4.71
predicted	4.62	4.50	4.65	4.71	4.95	5.05	5.18	5.28	5.18	5.20	5.10	5.02
high 95%	4.88	4.76	4.92	4.98	5.22	5.33	5.47	5.57	5.47	5.49	5.40	5.32

c) **Answer:** We present the following graph below, and note that the standard deviation in absolute difference between predicted values and realized is 0.1097166, which both seem to indicate a pretty good fit in the model.



Prediction Values and Limits for our SARIMA$(0,1,1) \times (0,1,1)_{12}$ Model												
high 95%	5.00	4.90	5.02	5.12	5.33	5.43	5.56	5.66	5.58	5.62	5.51	5.49
predicted	4.74	4.63	4.75	4.85	5.06	5.16	5.28	5.38	5.30	5.33	5.22	5.19
realized	4.54	4.49	4.77	4.62	5.01	5.01	5.25	5.37	5.15	5.14	5.16	4.81
low 95%	4.48	4.37	4.49	4.58	4.78	4.88	5.01	5.10	5.01	5.05	4.93	4.90

2.7 Seasonally Adjusting Data

Data on the number of monthly traffic fatalities in Ontario from 1960 to 1974 are contained in a file `fatalities.txt` on Blackboard. You may want to analyze the logs of the data.

(a) Use the function `stl` to seasonally adjust the data. (Some details on `stl` are available using the help facility in R – `help(stl)` – and more in the paper by Cleveland et al. on Blackboard.

The two key tuning parameters in `stl` are `s.window` and `t.window`, which control the number of observations used by loess in the estimation of the seasonal and trend components, respectively; these parameters must be odd numbers. For example;

```
> fatal <- scan(\fatalities.txt")
> fatal <- ts(fatal,start=c(1960,1),end=c(1974,12),freq=12)
> r <- stl(fatal,s.window = 3 , t.window = 51)
> plot(r)
> r <- stl(fatal,s.window = 5, t.window = 61)
> plot(r)
> r <- stl(fatal,s.window = \periodic", t.window = 41) # periodic seasonal comp
> plot(r)
```

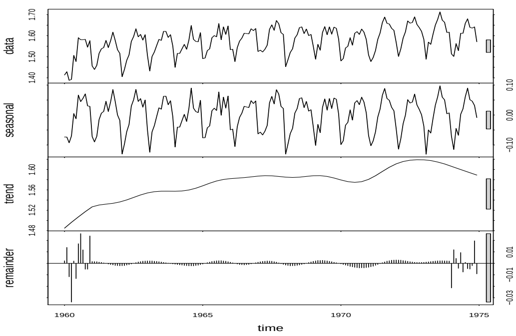
There is also an option `robust = T`, which allows one to better see anomalous observations or outliers in the irregular component.

(b) For one of set of parameter values used in part (a), look at the estimated irregular component. Does it look like white noise? Would you expect it to look like white noise?

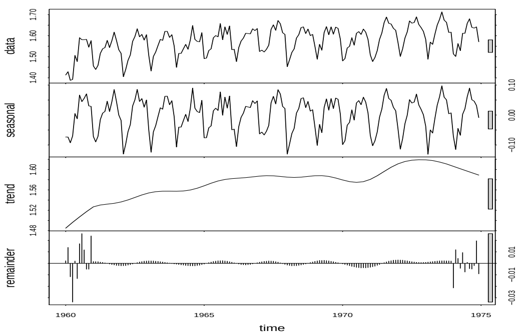
(c) The function `stl` estimates trend, seasonal, and irregular components. Other seasonal adjustment procedures can also estimate a calendar component in order to reflect variation due to the number of weekend days etc. For these data, do you think a calendar component would be useful?

(a) **Answer:** To do this question thoroughly, we created two loops to store, then plot all relevant seasonally adjusted data. Naturally since we have 456 plots, we do not print them all here. Instead, please visit https://github.com/jmostovoy/Time_Series/tree/master/A4_STL_Plots to view all plots. We also do the same process but with the robust setting turned on, and those plots may be viewed at https://github.com/jmostovoy/Time_Series/tree/master/A4_STL_Robust_Plots. We do however print a few of these plots here for illustrative purposes:

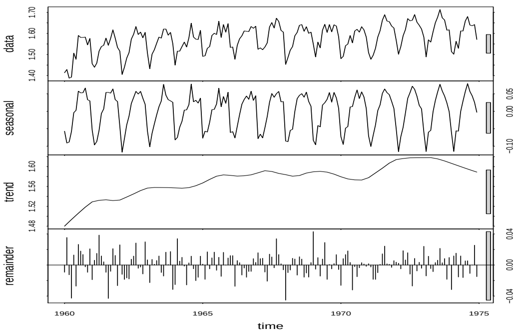
stl for t=23 and s=1



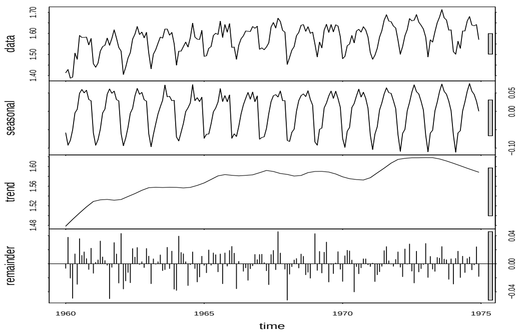
stl for t=23 and s=3



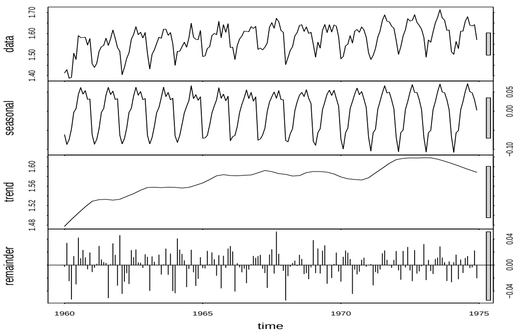
stl for t=23 and s=5



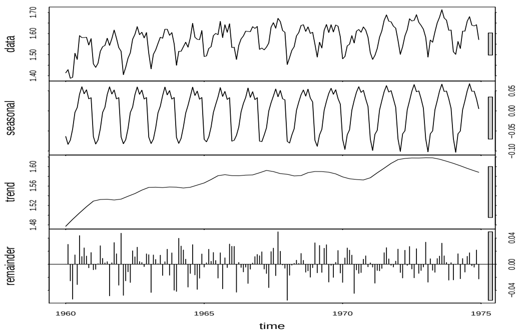
stl for t=23 and s=7



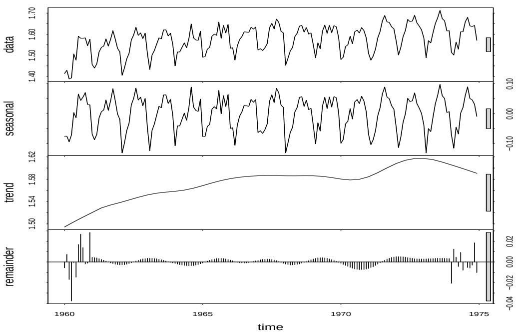
stl for t=23 and s=9



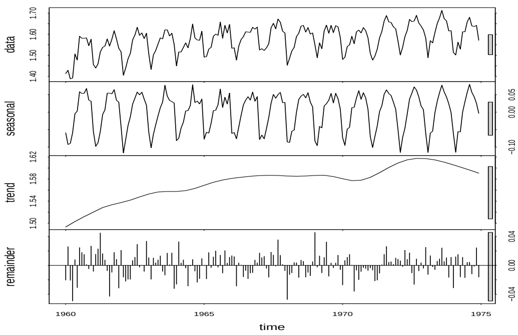
stl for t=23 and s=11



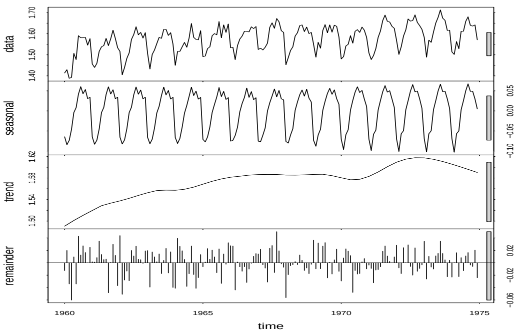
stl for t=35 and s=1



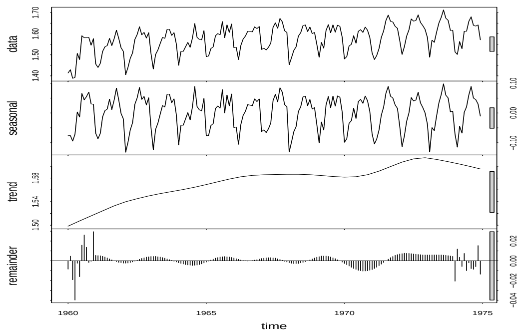
stl for t=35 and s=5



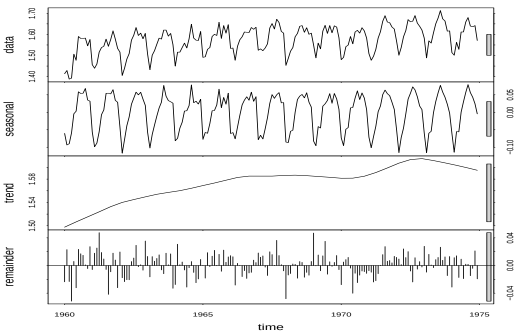
stl for t=35 and s=11



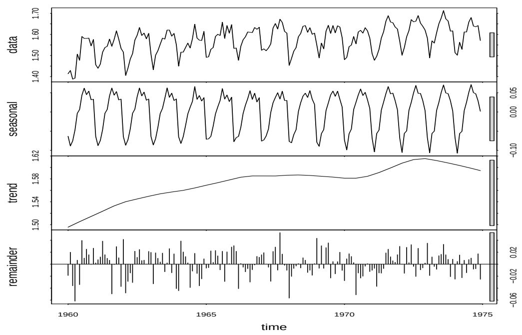
stl for t=47 and s=3



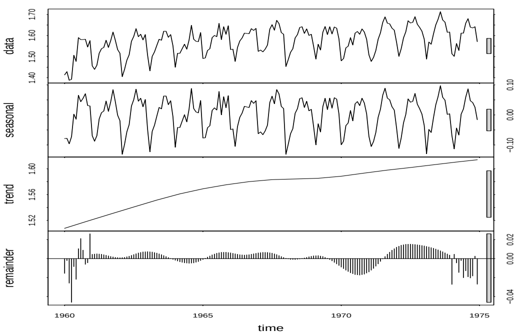
stl for t=47 and s=5



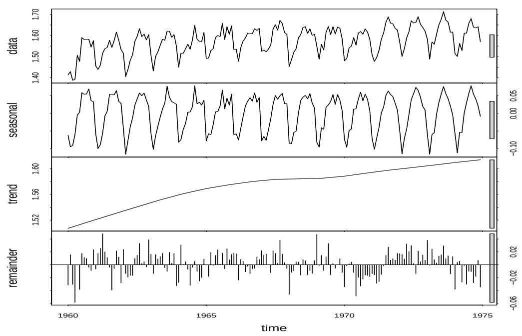
stl for t=47 and s=9



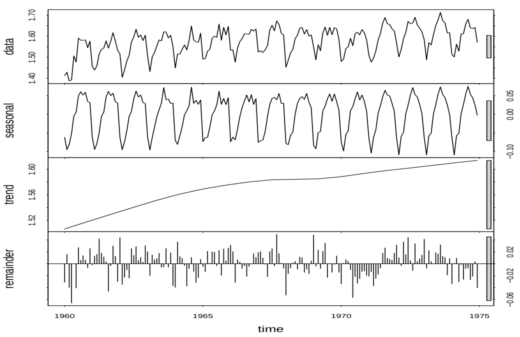
stl for t=95 and s=3



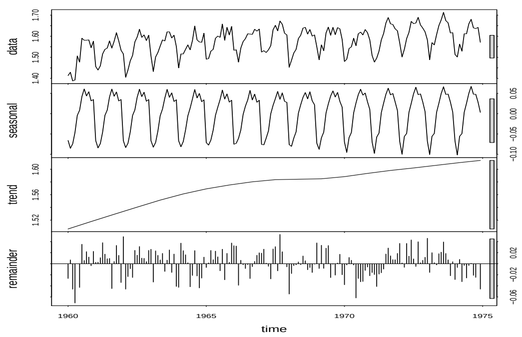
stl for t=95 and s=5

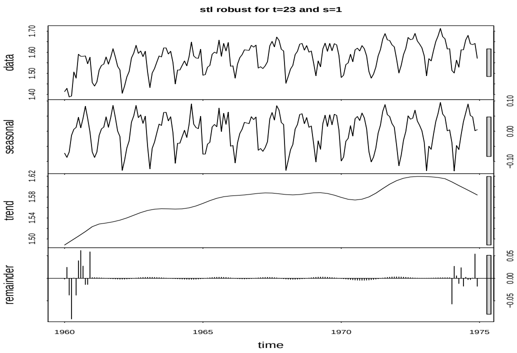


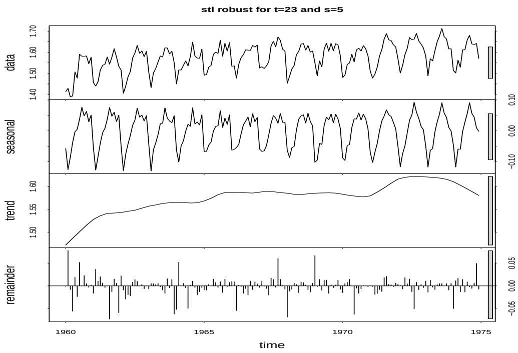
stl for t=95 and s=7

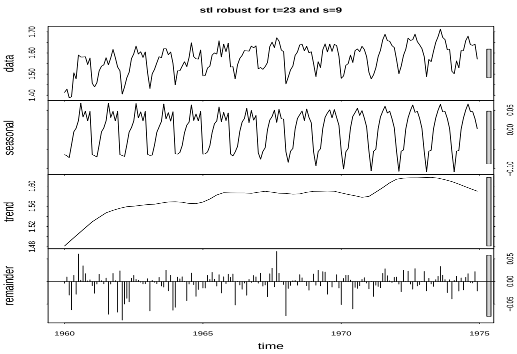


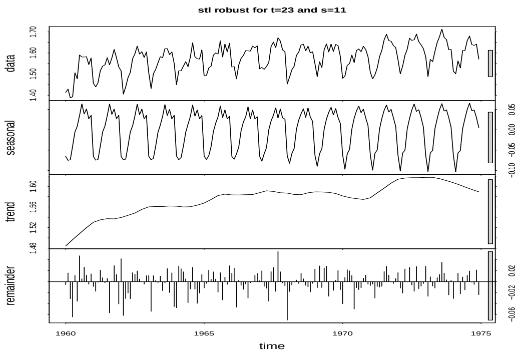
stl for t=95 and s=11

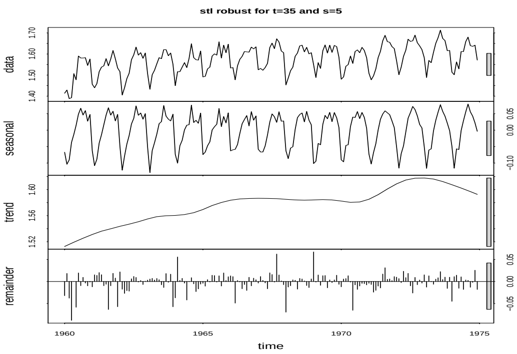


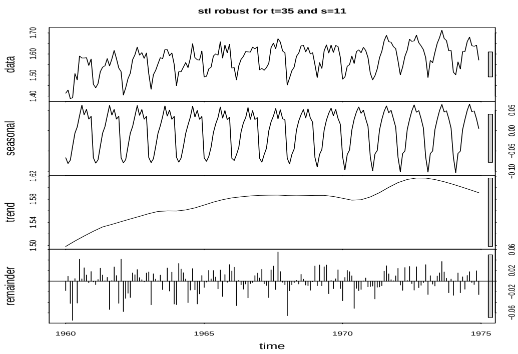

stl robust for t=23 and s=1
data
seasonal
trend
remainder
1960
1965
1970
1975
time

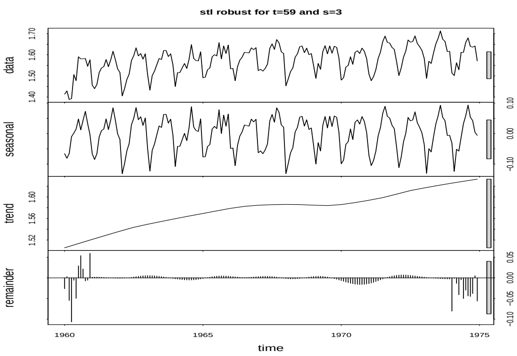

stl robust for t=23 and s=5
data
seasonal
trend
remainder
1960
1965
1970
1975
time

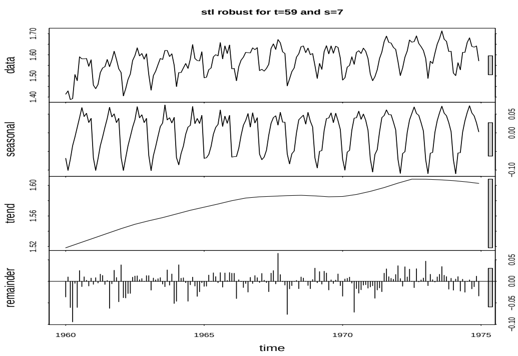

stl robust for t=23 and s=9
data
seasonal
trend
remainder
1960
1965
1970
1975
time


stl robust for t=23 and s=11
data
seasonal
trend
remainder
1960
1965
1970
1975
time


stl robust for t=35 and s=5
data
seasonal
trend
remainder
1960
1965
1970
1975
time


stl robust for t=35 and s=11
data
seasonal
trend
remainder
1960
1965
1970
1975
time

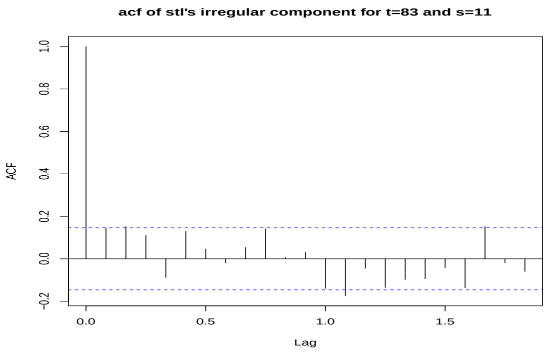

stl robust for t=59 and s=3
data
seasonal
trend
remainder
1960
1965
1970
1975
time

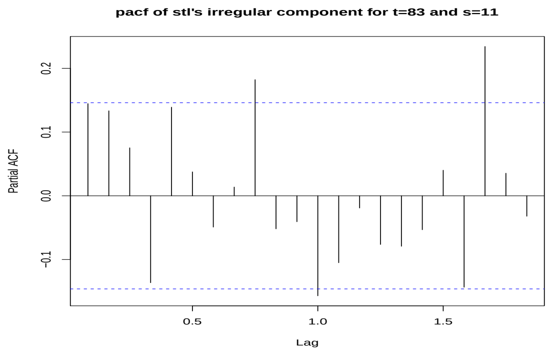

stl robust for t=59 and s=7
data
seasonal
trend
remainder
1960
1965
1970
1975
time

(b) **Answer:**

Just as in part (a), we sort of go overkill for what is being asked and plot ACFs and PACFs for all 456×2 data sets. These plots may be found in entirely at `https://github.com/jmostovoy/Time_Series/tree/master/A4_STL_ACFs` and `https://github.com/jmostovoy/Time_Series/tree/master/A4_STL_PACFs`. However, for an explicit example, we present one below:

Let us choose the `stl` data for $t = 83, s = 11$ as superficially it looks like it is both quite smooth and the residuals are about white noise. We look further into our suspicions by looking at both formal and informal tests for white noise. By plotting the ACF and PACF:





We can see that there for the ACF, there are (nearly) no significant lags for either plots, and the PACF isn't too bad for significant lags that we cannot yet reject the existence of white noise. Additionally, we have a p-value of 0.04262321 for the Bartlett Test, which again isn't tremendously significant either way. Furthermore, we can also see that the p-values for different lags of our box.test are presented below

	1	2	3	4	5	6	7	8	9	10
p-values	0.0503	0.0178	0.0160	0.0191	0.0108	0.0180	0.0315	0.0437	0.0195	0.0316

Thus, from these tests, we see that it the remainder part after fitting is likely not white noise, but we do not have quite enough significant data to very certainly refute the claim of white noise.

(c) **Answer:** Having a calendar component would certainly be useful. For example, certain attributes within the calendar year such as long holidays, I.e., Christmas & New Years vs. the rest of December would likely have the number of car fatalities dependent on these two different times within the same Month. Additionally, weather can vary quite dramatically from month to month, and within months. One plausible explanation for why there might a be greater number of accidents on holidays, long weekends is due to greater intoxication occurring at festivities during these times.

2.8 Spectral Densities

The file `speech.txt` contains a "speech record" of a person saying the syllable ahh; this was sampled at 10000 points per second. (These data represent a subset of a larger data set.)

(a) An R function, `spec.parzen`, is available for doing spectral density estimation using Parzen's lag window. For example;

```
> speech <- ts(scan(\speech.txt"),frequency=10000)
> r <- spec.parzen(speech,maxlag=60,plot=T)
```

will compute the estimate using M = 60; the plot will give approximate pointwise 95% confidence intervals for the spectral density function. Play around with different values of M to see how the estimates change with M. (Defining `speech` using ts with `frequency=10000` gives the frequency measured in Hertz or cycles per second.)

(b) Autoregressive spectral density estimates can be computed using `spec.ar`. For example,

```
> r <- spec.ar(speech,order=10,method=\burg")
```

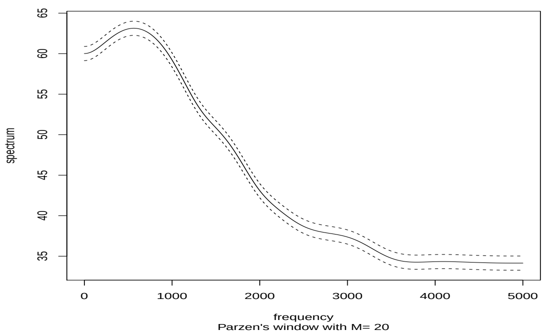
will give an estimate obtained by fitting an AR(10) model to the time series (using Burg's estimates) while

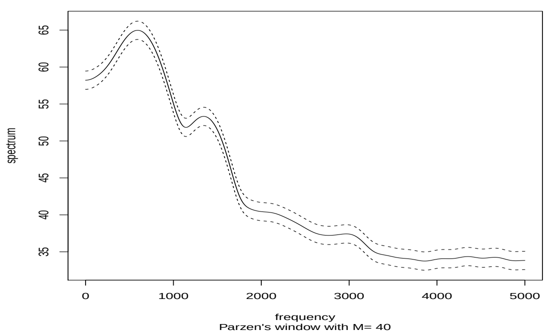
```
> r <- spec.ar(speech,method=\yw")
```

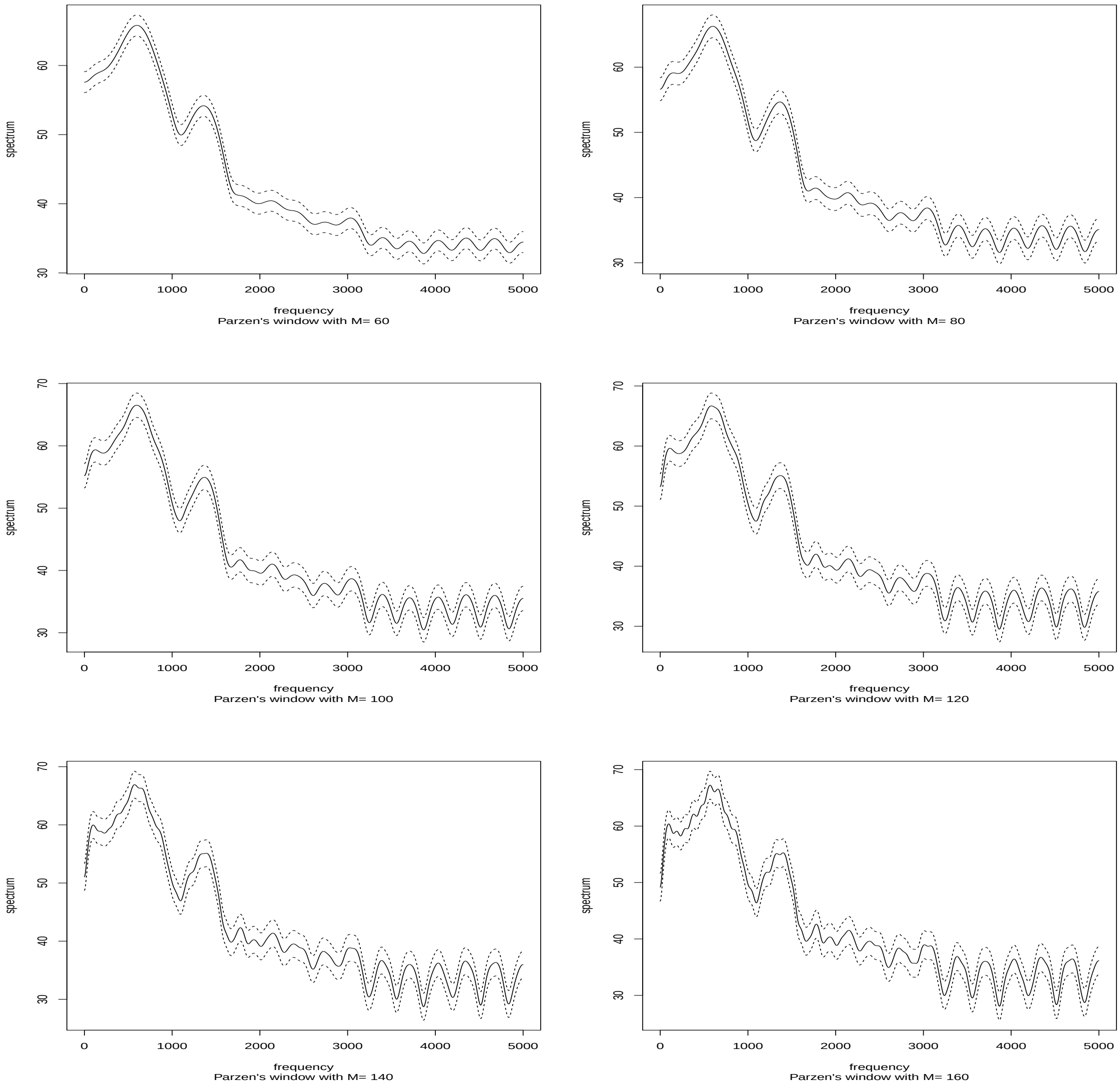
will use AIC to choose the AR order (using Yule-Walker estimates). Again play around with different AR orders and compare these to the estimates this estimate with the estimate obtained in part (a).

(c) Which frequencies (measured in Hertz or cycles per second) seem to be most dominant?

(a) **Answer:** We have stored all Spectral density plots using Parzen's lag window at the following link: https://github.com/jmostovoy/Time_Series/tree/master/A4_SP_Plots. For illustrative purposes, we show here 8 plots, increasing M by 20 for each plot:



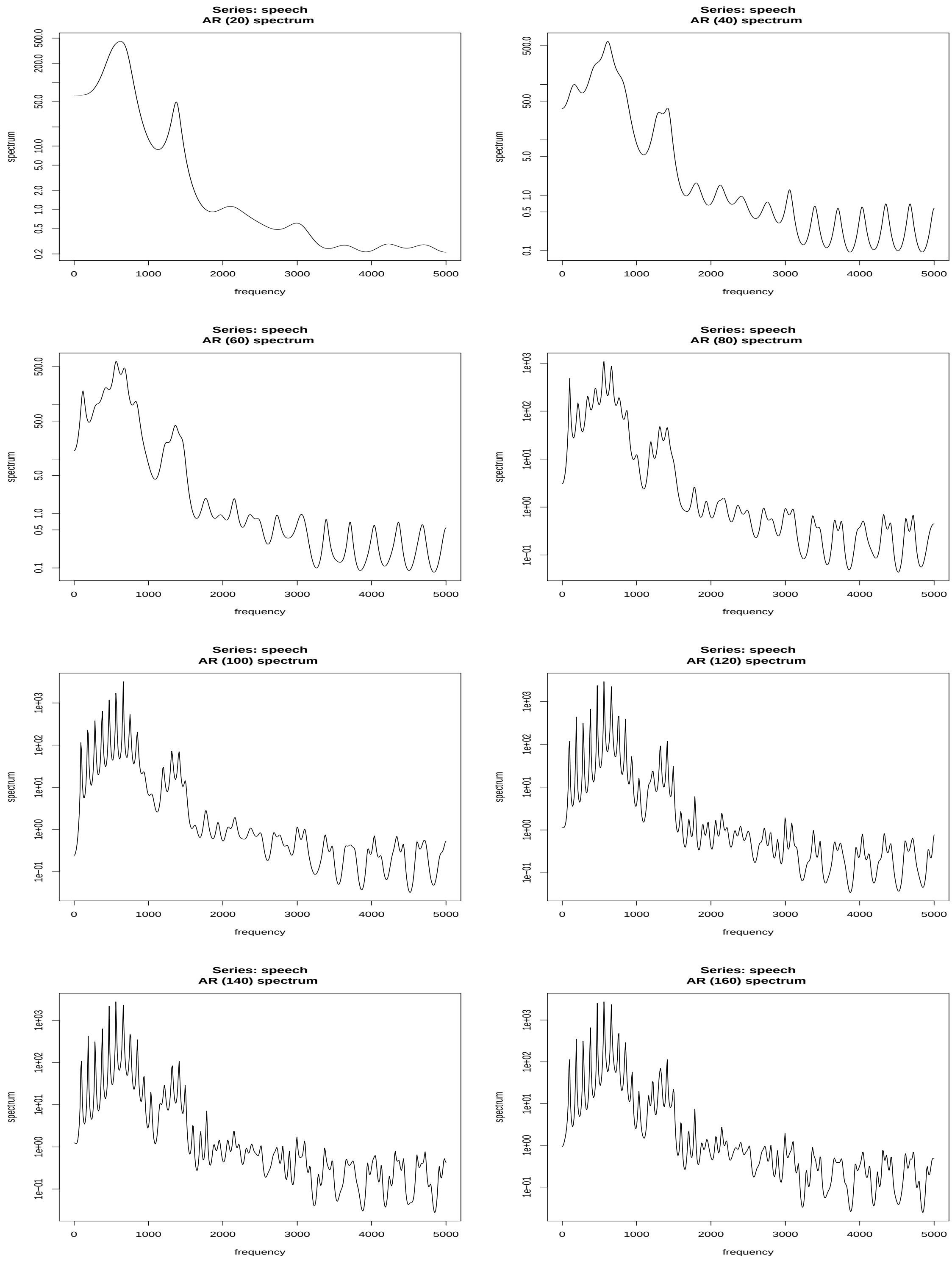




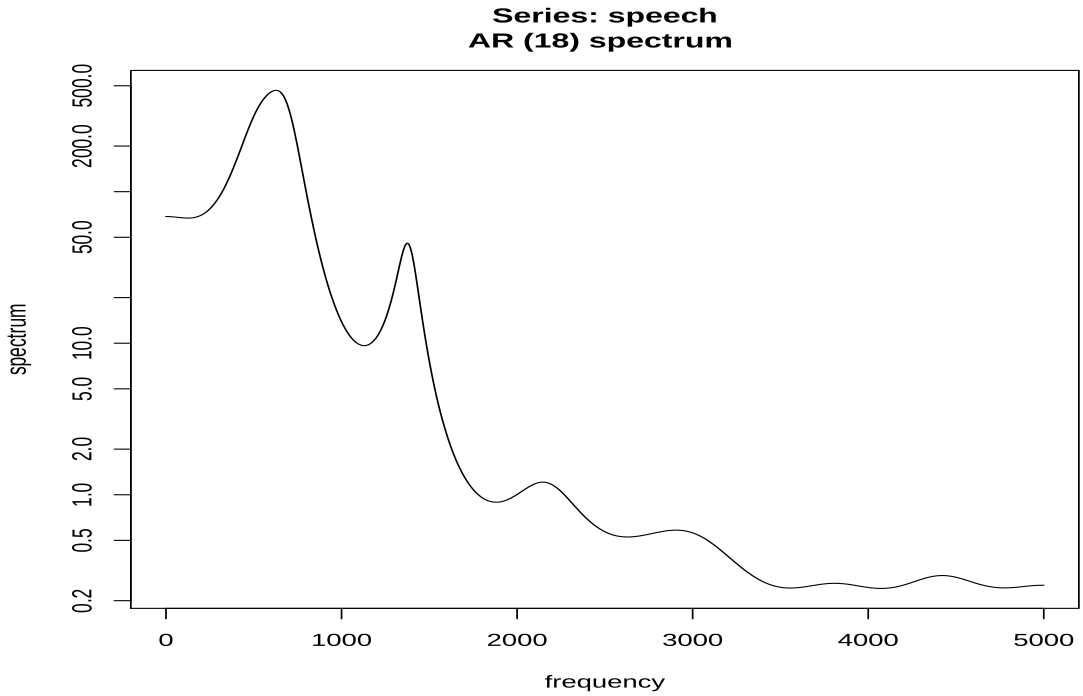
Thus, by noticing the trend occurring in the above plots as we increase M, we can see that the spectral density function appears smoother with a smaller M, and as M increases, there exists greater "ripples".

(b) **Answer:**

We present all plots for order $\in \{1, \ldots, 200\}$ at the following link: `https://github.com/jmostovoy/Time_Series/tree/master/A4_SA_Plots`. For illustrative purposes, we show here 8 plots, increasing the order by 20 for each plot:


Series: speech
AR (20) spectrum
spectrum
frequency
Series: speech
AR (40) spectrum
spectrum
frequency
Series: speech
AR (60) spectrum
spectrum
frequency
Series: speech
AR (80) spectrum
spectrum
frequency
Series: speech
AR (100) spectrum
spectrum
frequency
Series: speech
AR (120) spectrum
spectrum
frequency
Series: speech
AR (140) spectrum
spectrum
frequency
Series: speech
AR (160) spectrum
spectrum
frequency

Again, by noticing the trend occurring in the above plots as we increase the order, we can see that these plots appears smoother with a smaller orders, and as the order increases, there exists greater "ripples". If we use Yule-Walker estimates, we see that an order of 18 was chosen as optimal, and we present that plot below:



Furthermore, we can see that the dominant frequencies are only slightly different between the plots in (a) when comparing them to those in (b).

(c) **Answer:** The Yule Walker Estimate (order = 18) of the most dominant frequency of our Autoregressive spectral density is 631.2625. Furthermore, we can see that the dominant frequencies for order $= 20, 40, \ldots, 160$ is 621.2425, 611.2224, 571.1423, 561.1222, 661.3226, 561.1222, 561.1222, 561.1222 respectively.

2.9 Fitting an ARCH/GARCH Model

Daily stock prices (adjusted for stock splits and dividends) for Barrick Gold (from January 12, 1995 to November 14, 2016) are given in the file `barrick.txt`; the data are already transformed by taking logs and you should analyze them on this scale.

(a) Fit ARIMA(0,1,1) and ARIMA(0,1,2) models to the data. Do these models seem to fit the data adequately? Which model do you prefer and why? (You may want to do an ADF test to see if the time series needs to be differenced to make it stationary but the non-stationary should be quite clear.)

(b) Using the residuals from your preferred model from part (a), fit ARCH(m) models for $m = 1, 2, 3, 4, 5$. Which model seems to be the best?

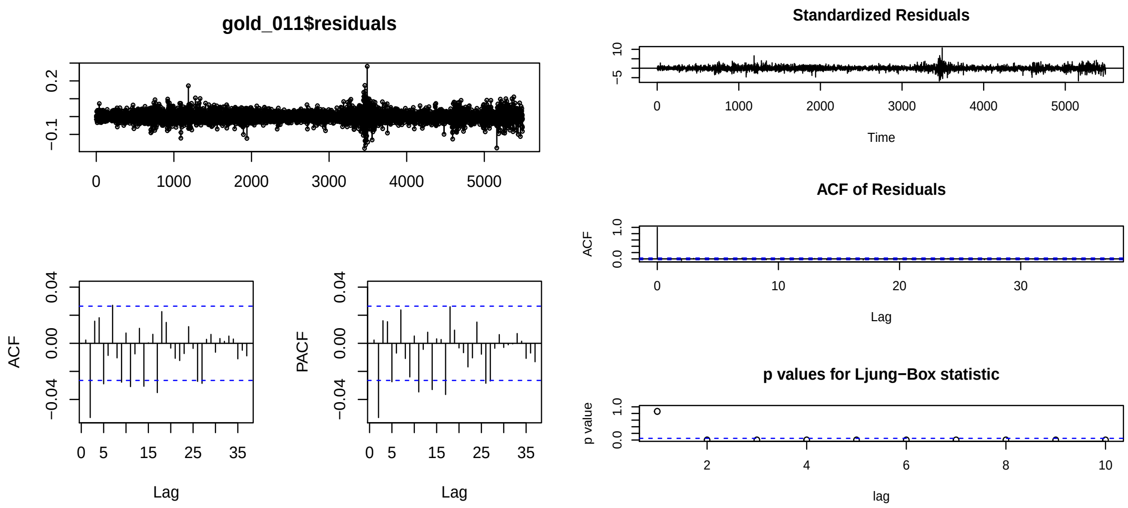
(c) Repeat part (b), using GARCH$(1, s)$ models for $s = 1, 2, 3, 4, 5$. Are any of these models an improvement over the best ARCH model from part (b)?

Note: The R package `fGarch` contains the function `garchFit`, which can be used to fit both ARCH and GARCH models. Some more information on `garchFit` will be provided on Blackboard.

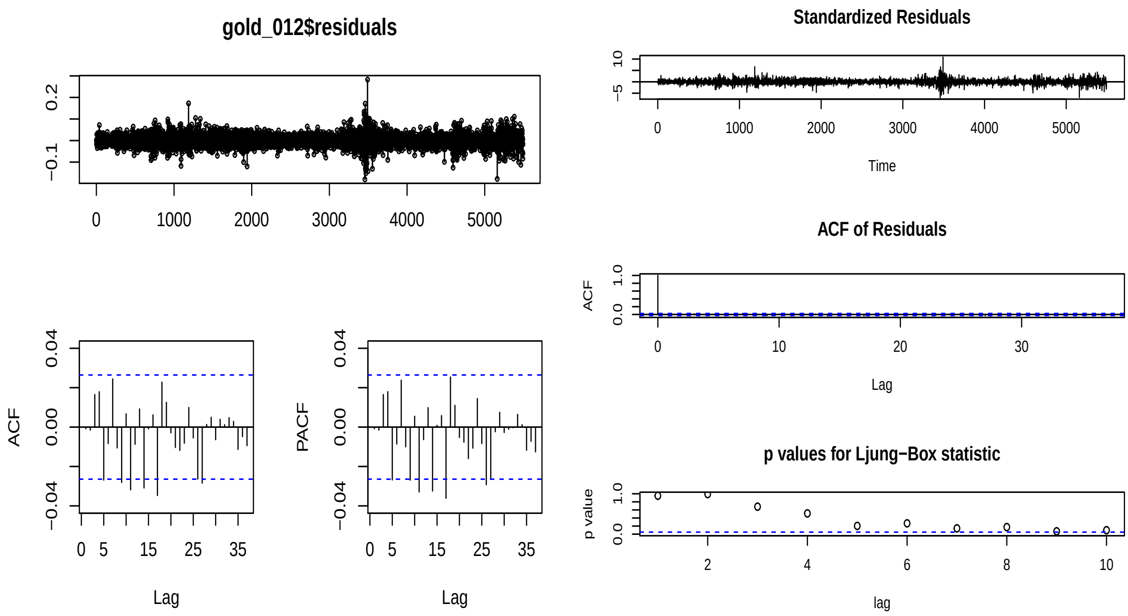
1. **Answer:** The ADF for the non-differenced and the differenced data have p-values of 0.3879 and $x < 0.01$ respectively, and hence we certainly should use the difference the data.

   Furthermore, we can see quite clearly see that from different types of tests, we prefer an ARIMA $(0, 1, 2)$ model over the ARIMA$(0, 1, 1)$. The AIC for the ARIMA$(0, 1, 2)$ model and the ARIMA $(0, 1, 1)$ model are -24450.7 and -24438.09 respectively (the ARIMA$(0, 1, 2)$ has a lower AIC and hence preferable). The Bartlett p-values on the residuals from the ARIMA$(0, 1, 2)$ model and the ARIMA$(0, 1, 1)$ model are 0.4142057 and 0.004693974 respectively, and hence the ARIMA$(0, 1, 2)$ again beats out the ARIMA$(0, 1, 1)$. Furthermore, we can see from the following plots that the ARIMA$(0, 1, 2)$ model is far more superior than the ARIMA$(0, 1, 1)$ model:

ARIMA$(0, 1, 1)$



ARIMA$(0, 1, 2)$



2. **Answer:** Based on the AICs of the ARCH(m), $m = 1, \ldots, 5$ models using the residuals from ARIMA(0,1,2), we find that the ARCH(5) model has the lowest AIC of -4.588344 indicating the best fit. The AICs for the ARCH(m), $m = 1, \ldots, 5$ is -4.503860, -4.533235, -4.562443, -4.576758, -4.588344 respectively.

   This is also confirmed by looking at the BICs, which are also decreasing with greater m: -4.500249, -4.528421, -4.556425, -4.569536, -4.579919 and hence also indicate ARCH(5) is preferable.

3. **Answer:** Based on the AICs of the GARCH$(1, q)$, $q = 1, \ldots, 5$, we see that the GARCH$(1, 1)$ is our best fit, and also has a slightly better AIC when compared to ARCH(5). The AICs for GARCH$(1, q)$, $q = 1, \ldots, 5$ are -4.664868, -4.664444, -4.664042, -4.663623, -4.663207 respectively.

   This is also confirmed by looking at the BICs, which are also increasing with greater q: -4.660054, -4.658426, -4.656821, -4.655198, -4.653578, and hence also indicate GARCH$(1, 1)$ is preferable.