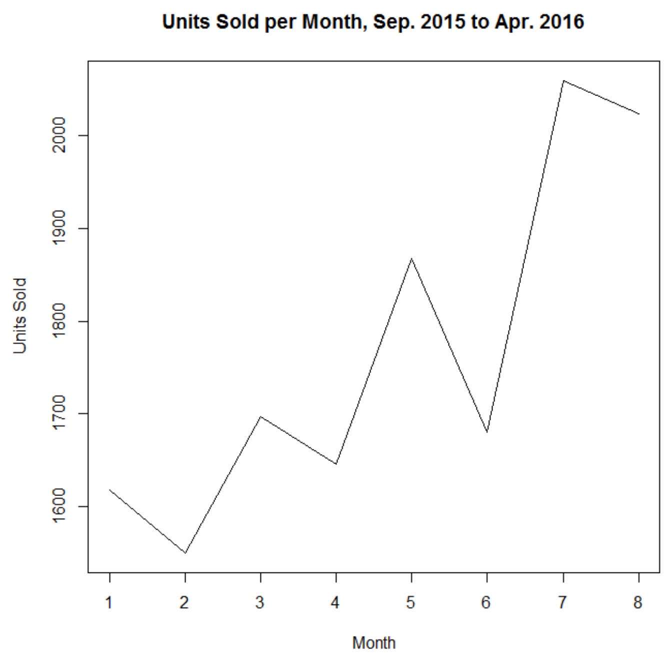
In the following, B denotes the backward shift operator and $Z(t)$ denotes a white noise process with mean zero and variance σ^2 unless otherwise stated. Where questions ask you to explain or justify your thinking, answers not written in sentences may be penalised. For questions 1 – 10 please circle your answers below. Ambiguous responses will be considered incorrect.

1. (a) (b) (c) (d) (e)
2. (a) (b) (c) (d) (e)
3. (a) (b) (c) (d) (e)
4. (a) (b) (c) (d) (e)
5. (a) (b) (c) (d) (e)
6. (a) (b) (c) (d) (e)
7. (a) (b) (c) (d) (e)
8. (a) (b) (c) (d) (e)
9. (a) (b) (c) (d) (e)
10. (a) (b) (c) (d) (e)

1. After completing this course, you are offered a lucrative consulting job for a company that manufactures electrical relays that are used as components in certain makes of cars. The relay went into production quite recently, and there are only eight months of sales figures. The time series plot is below, showing the number of units sold per month.



The company want you to forecast the next three months of unit sales for their relay. Which of the following would be a sensible initial response? (Circle all that apply.)

(a) Fit a SARIMA model to the data and use Box–Jenkins forecasting.

(b) Fit a Gompertz curve to the data, and extrapolate to future time points to give predictions.

(c) Compute the periodogram for the data.

(d) Smooth the data using a centred moving average filter, compute seasonal indices, and base your predictions on the sample mean and the seasonal indices.

(e) **Ask the company some questions about their relay production and sales, such as why there has been an apparent**

upward trend, why sales in some months are lower than in the previous month, and whether new consumers of their product are expected.

2. Consider the model

$$X(t) = 0.8Z(t-2) + 0.9Z(t-1) + Z(t).$$

The stochastic process $X(t)$ is (Circle all that apply.)

(a) neither stationary nor invertible.

(b) stationary but not invertible.

(c) invertible but not stationary.

(d) **both stationary and invertible.**

(e) a random walk model.

The process $X(t)$ is stationary by definition of $Z(t)$, and is invertible if the roots of

$$\theta(B) = 1 + 0.9B + 0.8B^2$$

are outside the unit circle. To check this, the roots are

$$\frac{-0.9 - \sqrt{0.9^2 - 4 \times 0.8}}{2 \times 0.8},$$

that is, $-0.562 \pm 0.966\,2i$*. These roots have absolute value* $\sqrt{0.562^{\,2} + 0.966\,2^2} = 1.117\,8 > 1$*, so* $X(t)$ *is invertible.*

3. Consider the model

$$X(t) = \alpha + \beta t + Z(t)$$

for some constants α and β. Then $X(t)$ is

(a) not an ARIMA model.

(b) an AR(1) process.

(c) an MA(1) process.

(d) an ARIMA(1, 1, 0) process.

(e) **an ARIMA**$(0,1,1)$.

4. Let c_k be the sample autocovariance function at lag k for a time series $x(1), \ldots, x(N)$. Which of the following are properties of c_k? (Circle all that apply.)

   (a) $c_0 = 1$.

   (b) $\mathbf{c}_{-k} = \mathbf{c}_k$.

   (c) c_k is independent of c_{k-1}.

   (d) c_k decreases as k increases.

   (e) **If the data exhibit an upward trend there will be a trend in c_k, for $k = 1, 2, \ldots, N/4$.**

5. Consider the stochastic process

$$X(t) = X(t-1) + \alpha X(t-2) - \alpha X(t-3) + Z(t).$$

   The process $X(t)$ is stationary

   (a) **for no values of α.**

   (b) for all values of α.

   (c) when $|\alpha| < 1$.

   (d) when $-\frac{1}{2} < \alpha < \frac{1}{2}$.

   (e) when $\alpha > 0$.

   The stationarity is determined by the roots of

$$1 - B - \alpha B^2 + \alpha B^3.$$

   But this has a unit root $B = 1$ so $X(t)$ is not stationary for any value of α.

6. A stock broker is interested in forecasting the price of a certain stock (in \$). This figure is recorded at the close of trading on each of two hundred days, and a statistician decides a good fit to the series is the model

$$X(t) - 2.41 = 0.87(X(t-1) - 2.41) + Z(t).$$

   The most recent closing price was \$2.81. If forecasting the value of $X(t)$ at lead time l using the model above, as l increases the forecast value will converge to

(a) zero.

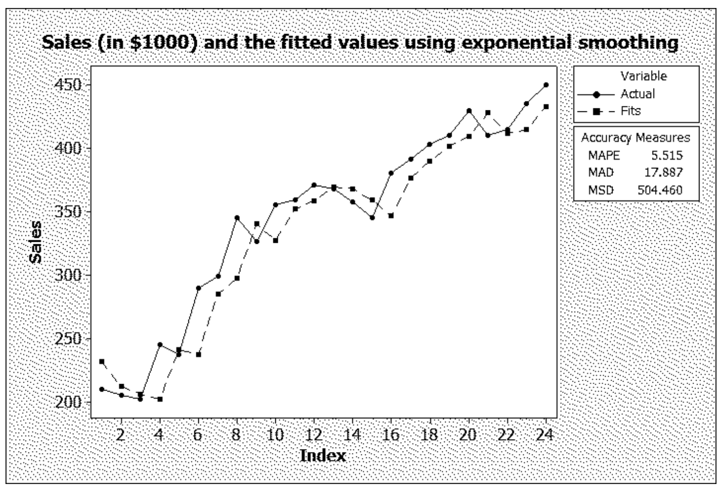
(b) \$0.13.

(c) \$2.10.

(d) **\$2.41.**

(e) \$2.81.

7. The following is a plot of the monthly sales figures (in \$000) for a company over a two–year period. Exponential smoothing was applied to the series, and the fitted values are also included on the plot below:



Based on the above, select the most likely value of the smoothing parameter α that was applied:

(a) –0.1

(b) 0.1

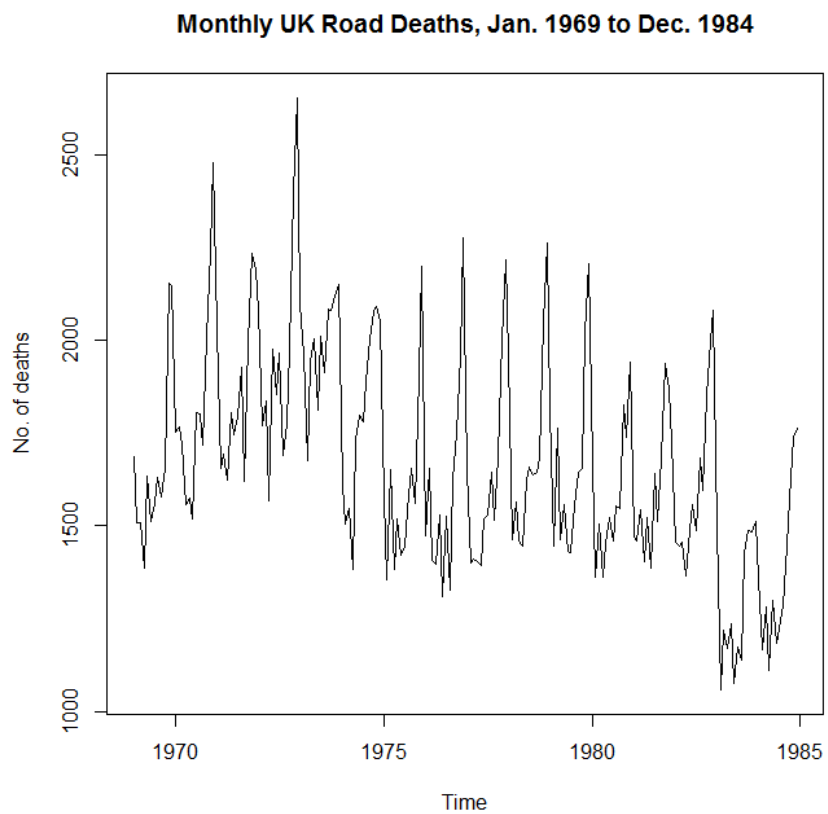
(c) 0.4

(d) *0.9*

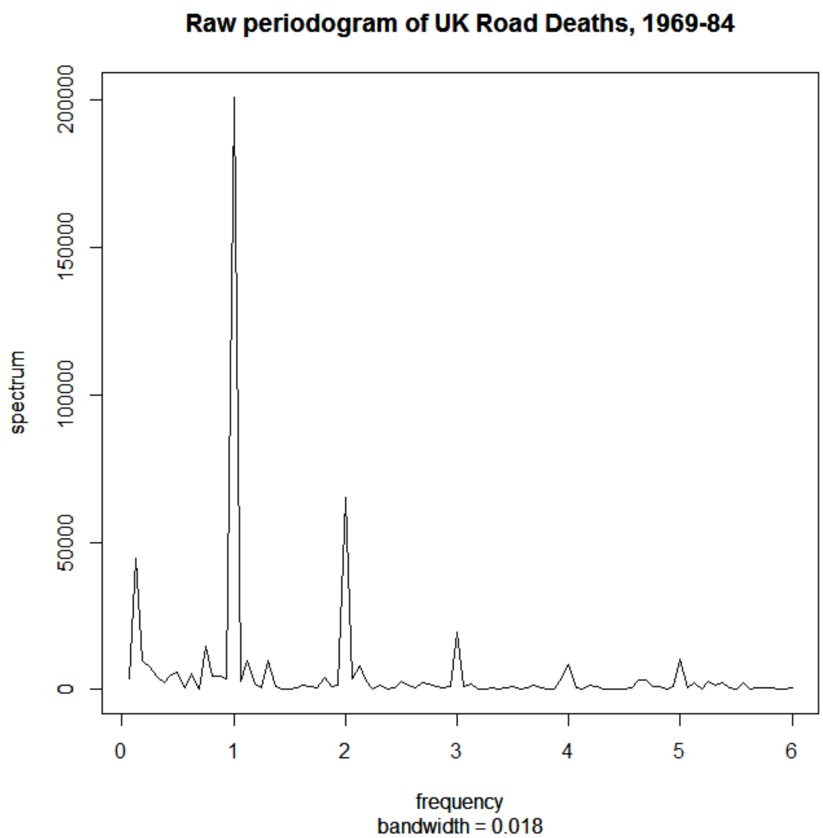
(e) 1.5

8. The time series plot below shows the monthly UK deaths due to road

accidents, January 1969 to December 1984.



Below is a plot of the unlogged periodogram of the time series, where unit time is taken as one year.



There is a clear spike in the periodogram at the frequency labelled "1" above. Counting the frequencies at which the periodogram above is evaluated from the left, the frequency with the largest periodogram value would be the

(a) 1st
(b) 6th
(c) 8th
(d) 12th
(e) **16th**
There are 192 observations in total. The frequency scale is

$$f = \frac{p}{192\delta t}$$

where $\delta t = 1/12$, and this is unity when $p = 16$.

9. We note there are smaller significant spikes in the periodogram at frequencies above the location of the major spike. This is most likely attributable to

   (a) noise in the periodogram.

   (b) *leakage*, in that the contributions from important frequencies at which the periodogram is not computed have been distributed to neighbouring frequencies.

   (c) **harmonics of the major frequency, due to the periodic effect being not exactly sinusoidal at the frequency with the largest contribution.**

   (d) *aliasing*, with frequencies outside the observable range being represented in the periodogram.

   (e) the periodogram being over–smoothed.

10. In the periodogram in Q.8, we note that the bandwidth is given to be 0.018. In general this number (circle all that apply)

    (a) **is a measure of the width of the spectral window (that is, the smoothing function that has been applied to the periodogram).**

    (b) would decrease if we increased the span of the smoothing applied to the periodogram.

    (c) is determined by the tapering parameter used when padding the data.

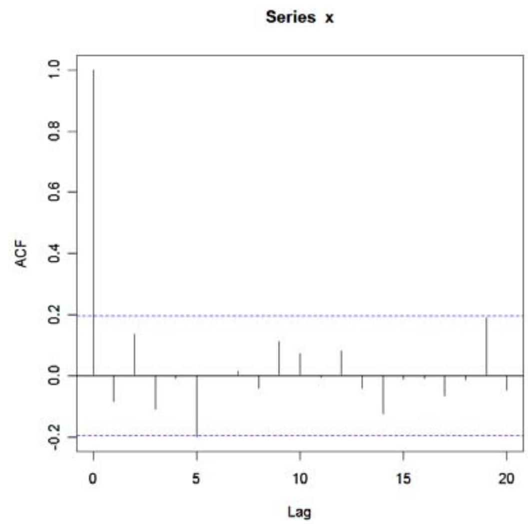
    (d) is proportional to the variance of the modified periodogram at a given frequency.

    (e) **is inversely proportional to the variance of the modified periodogram at a given frequency.**

11. Let $Z(t)$ be independent standard Normal variables, for $t = 1, 2, \ldots$. Describe, with the aid of sketches, how you would expect the correlogram to appear if based on a sample of size 100 taken from each of the following processes:

(a) $Z(t)$

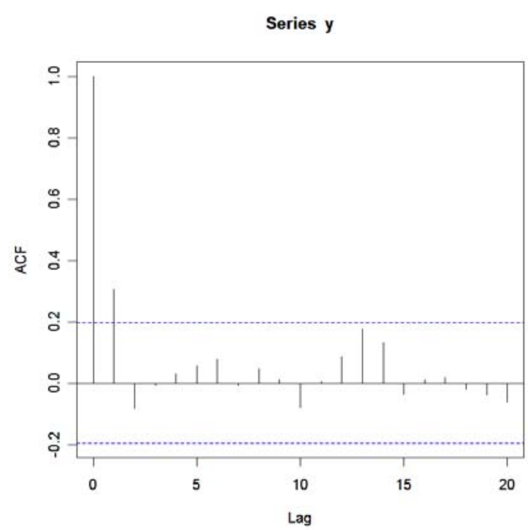
There should be no "significant" lags (outside ± 0.2) aside from lag 0, other than the occasional one by chance.



(1 mark)

(b) $X(t) = 0.3X(t-1) + Z(t)$

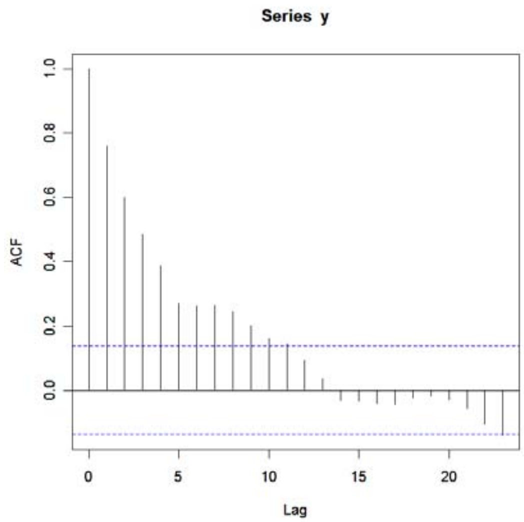
The value at lag 1 should be approximately 0.3, behaviour generally like $(0.3)^k$.



(2 marks)

(c) $X(t) = 0.8X(t-1) + Z(t)$

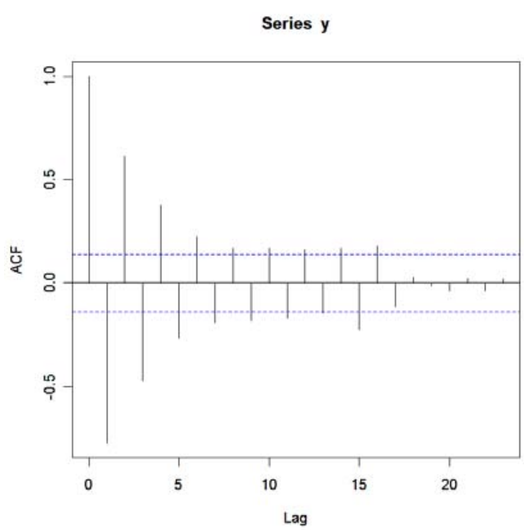
There should be a gradual tailing off – slow decay, like $(0.8)^k$.



(2 marks)

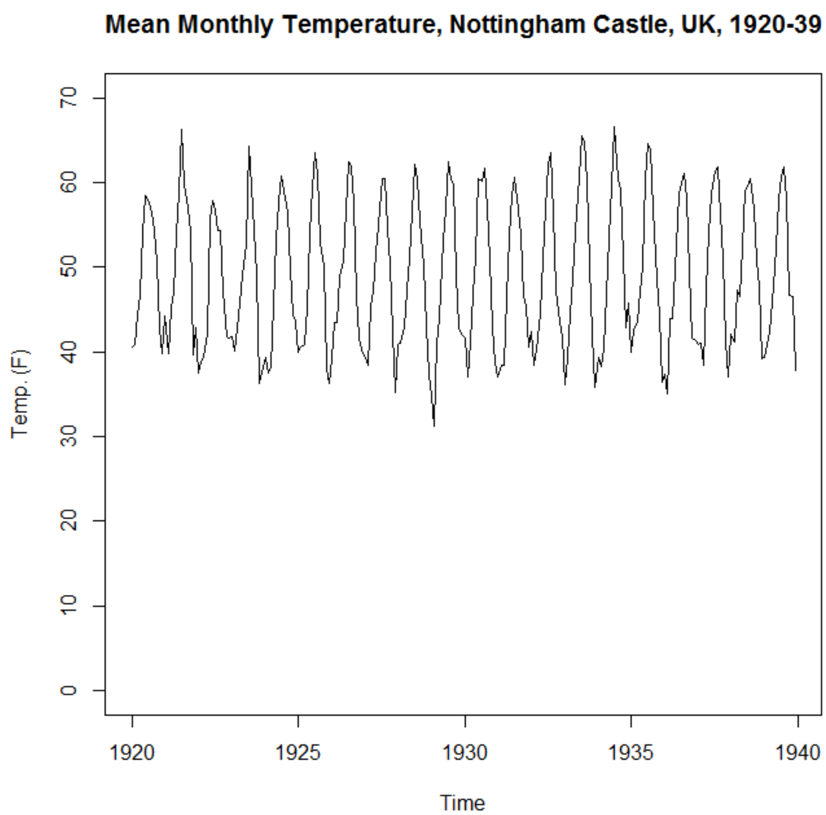
(d) $X(t) = -0.8X(t-1) + Z(t)$

Acf oscillates here, and decays quite slowly to 0.



(2 marks)

12. This question concerns forecasting methods. The time series plot shown below is for the average monthly air temperature (°F) at Nottingham Castle, UK, 1920–1939.



(a) Although all measurements are correct if identifying a possible outlier in the above time series, in which year would you suggest the possible outlier was recorded?
The February 1929 observation is a possible outlier. (1 mark)

(b) If applying Holt–Winters method to these data, would you select a multiplicative or additive model? Justify your reasoning.
An additive model seems best, as the amplitude of the seasonal effect seems to be relatively constant. (2 marks)

(c) The estimates of the seasonal effects for January and February 1939 are –10.524 and –10.638 (°F) respectively, and the December 1939 estimates of the level and trend components were 50.288 and –0.00124 respectively. What are the Holt–Winters forecast values for the monthly mean temperatures in January and February 1940?

The time t forecast for the value at time $t+l$ is then

$$\hat{x}(t,l) = L(t) + lT(t) + I(t-12+l).$$

which for January 1940 is

$$\begin{aligned}\hat{x}(1939.12,1) &= 50.288 - 0.00124 - 10.524 \\ &= 39.763\end{aligned}$$

Fahrenheit, and for February 1940,

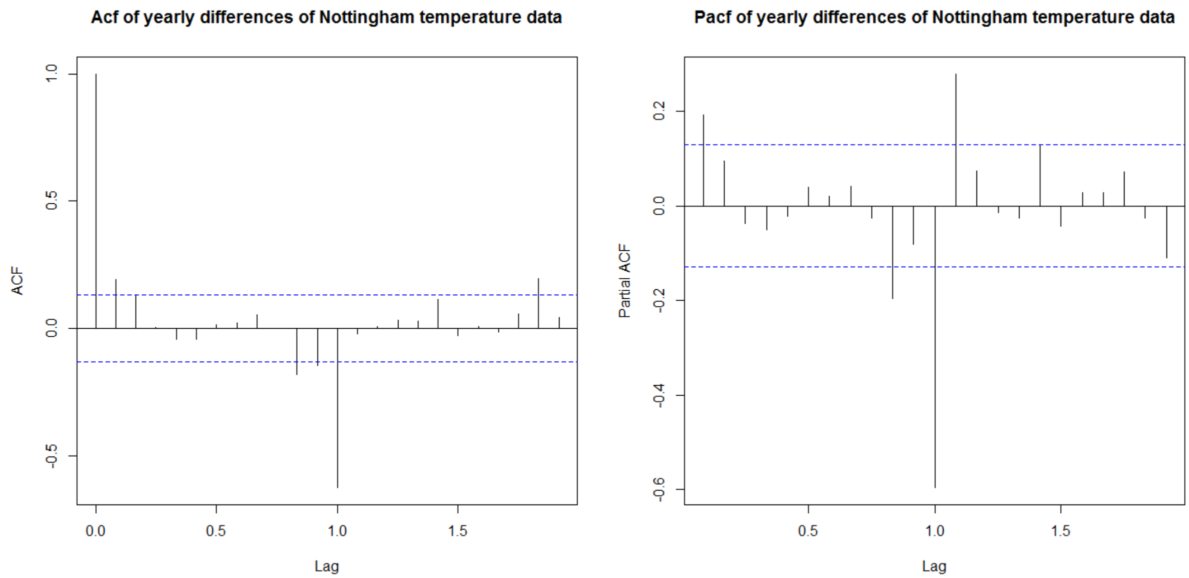
$$\begin{aligned}\hat{x}(1939.12,2) &= 50.288 - 2\times 0.00124 - 10.638 \\ &= 39.648\end{aligned}$$

F. (4 marks, 2 each. If a multiplicative model adopted, answers are of form

$$\hat{x}(1939.12,l) = (50.288 + l(-0.00124))\, I(t-12+l),$$

no additional marks deducted.)

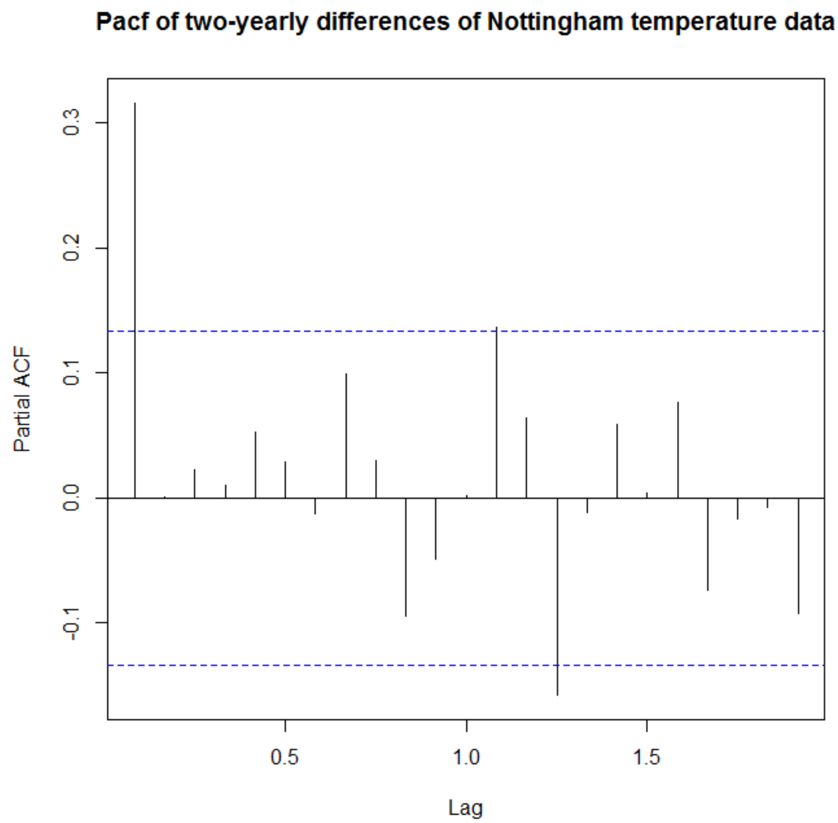
(d) If applying Box–Jenkins forecasting methods here it is appropriate to consider differencing the time series at lag 12, that is, look at yearly differences. The acf and pacf plots of the yearly differences are given below.



Comment on the above plots, with reference to possible models to fit to the data.
The acf plot is consistent with a AR model, apart from the large negative spike at lag 12. The pacf may cut off at lag 1, but has large absolute values at lags 12 and 13. These suggest a possible AR model with a non–seasonal and a seasonal term. (2 marks: 1 for comments plots, 1 for conclusion about possible model.)

(e) Describe how you think the pacf would look for the two-yearly differenced data (that is, the data differenced at lag 24)?
The pacf plot above suggests the pacf for the two–yearly differences

would cut off at lag 1. In fact this is observed:



(1 mark)

(f) A possible model is of the form

$$\left(1+0.860B^{12}+0.296B^{24}\right)(1-0.286B)\left(1-B^{12}\right)X(t)=Z(t)\,.$$

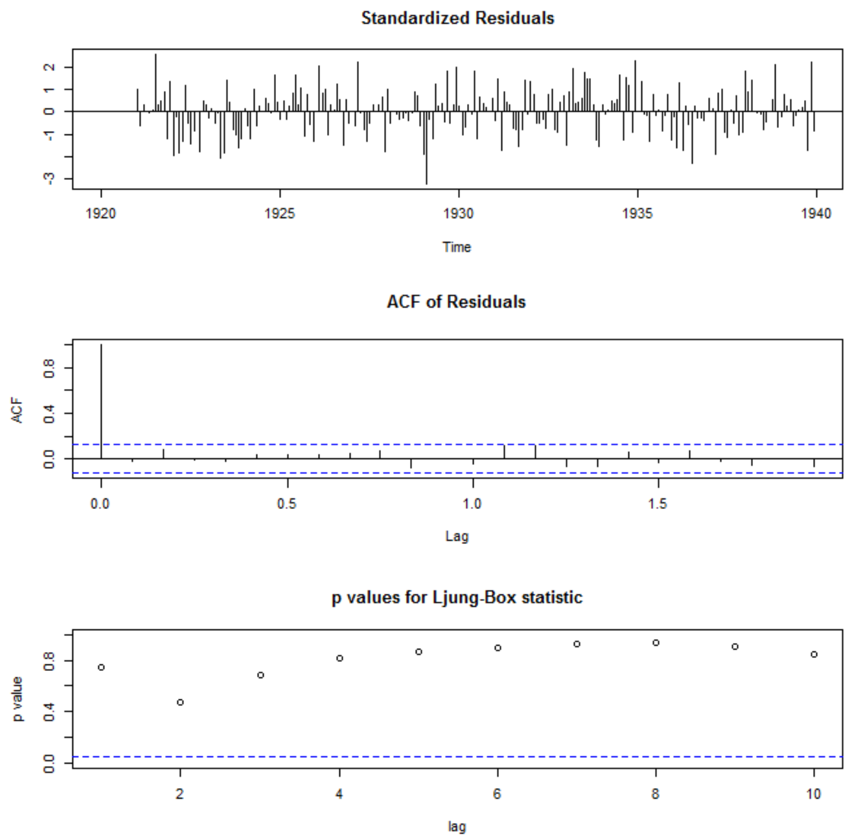
Identify the order of this SARIMA model.

We have taken differences at lag 12, implying $D=1$; *a polynomial of order* $P=2$ *in* B^{12} *acts alongside a linear function of* B. *This tells us the model is of order* $(1,0,0)\times(2,1,0)_{12}$. (2 marks)

(g) If written as an ARMA model, in which family does the model in (f) belong, and what is the order?

The model is an $AR(37)$ *model.* (2 marks, 1 for AR, 1 for order)

Group version: Below are diagnostic plots for the model fitted in (f). Comment on these plots and what they suggest about how well the model fits.



The standardised residuals are small in absolute value (even for the possible outlier), and the acf of the residuals does not suggest lingering autocorrelation. The p-values of the Ljung–Box statistics are all large for all lags. (2 marks: any two of the three main points acceptable.)

(h) If using the model in (f) to forecast the January 1940 value of the time series by hand (that is, not using software) given the complete time series, explain clearly the steps you would be required to undertake.

First, the brackets in the model in (f) would been expanded to give *the AR(37) form. Then at time $N+1$ the model is*

$$X(N+1) = \pi_1 X(N-36) + \pi_2 X(N-35) + \cdots + \pi_p X(N) + Z(N+1)$$

for some constants $\pi_1, \ldots, \pi_p$. For forecasting at time 240+1, we would simply plug in the observed values required up to time 240, and set $Z(N+1) = 0$. (2 marks: 1 for noting expansion of the model into AR form is required.)

13. This question is about the discrete–time stochastic process $X(t)$ defined as

$$X(t) = \cos(\pi t/3 + \theta),$$

where $\theta \sim U(0, 2\pi)$, and θ is fixed for a given realisation.

(a) Show clearly why $E(X(t)) = 0$.
Now

$$\begin{aligned} E(X(t)) &= E(\cos(\pi t/3)\cos(\theta) - \sin(\pi t/3)\sin(\theta)) \\ &= \cos(\pi t/3)\frac{1}{2\pi}\int_0^{2\pi}\cos(\theta)\,d\theta - \sin(\pi t/3)\frac{1}{2\pi}\int_0^{2\pi}\sin(\theta)\,d\theta \\ &= 0 \end{aligned}$$

(2 marks)

(b) Given that $\int_0^{2\pi}\cos^2(\theta)\,d\theta = \pi$, find $\mathrm{Var}(X(t))$, showing your working clearly.
We have $\int_0^{2\pi}\sin^2(\theta)\,d\theta = \int_0^{2\pi}(1-\cos^2(\theta))\,d\theta = \pi$, and so $\mathrm{Var}(X(t))$ *is*

$$\begin{aligned} E\left(X(t)^2\right) &= E(\cos^2(\pi t/3)\cos^2(\theta) + \sin^2(\pi t/3)\sin^2(\theta) \\ &\quad -2\cos(\pi t/3)\cos(\theta)\sin(\pi t/3)\sin(\theta)) \\ &= \frac{\cos^2(\pi t/3)}{2\pi}\int_0^{2\pi}\cos^2(\theta)\,d\theta + \frac{\sin^2(\pi t/3)}{2\pi}\int_0^{2\pi}\sin^2(\theta)\,d\theta \\ &\quad -\frac{\cos(\pi t/3)\sin(\pi t/3)}{\pi}\int_0^{2\pi}\sin(\theta)\cos(\theta)\,d\theta \\ &= \frac{1}{2}\left(\cos^2(\pi t/3) + \sin^2(\pi t/3)\right) \\ &= \frac{1}{2} \end{aligned}$$

since

$$\begin{aligned}\int_0^{2\pi} \sin(\theta)\cos(\theta)\,d\theta &= \frac{1}{2}\int_0^{2\pi}\sin(2\theta)\,d\theta \\ &= 0.\end{aligned}$$

(4 marks)

(c) Hence sketch the spectral distribution function of $X(t)$, for frequencies $\omega \in (0,\pi)$.
As all the power in the spectrum is at $\pi/3$ *we have*

$$F(\omega) = \begin{cases} 0 & \omega < \pi/3 \\ \\ \frac{1}{2} & \omega \geq \pi/3 \end{cases}$$

which jumps (and so is non-differentiable) at $\omega = \pi/3$. (2 marks)

(d) Find $\mathrm{Cov}(X(t), X(t+k))$.
As $\cos(A+B) + \cos(A-B) = 2\cos(A)\cos(B)$, we have

$$\begin{aligned} Cov(X(t), X(t+k)) &= E(\cos(\pi t/3 + \theta)\cos(\pi/3(t+k)+\theta)) \\ &= \frac{1}{2}E(\cos(2\pi t/3 + \pi k/3 + 2\theta) + \cos(\pi k/3)) \\ &= \frac{\cos(\pi k/3)}{2} + \frac{1}{2}E(\cos(2\pi t/3 + \pi k/3)\cos(2\theta) \\ &\quad - \sin(2\pi t/3 + \pi k/3)\sin(2\theta)) \\ &= \frac{\cos(\pi k/3)}{2} + \frac{1}{2}\left(\frac{\cos(2\pi t/3 + \pi k/3)}{4\pi}\int_0^{4\pi}\cos(x)\,dx\right. \\ &\quad \left. - \frac{\sin(2\pi t/3 + \pi k/3)}{4\pi}\int_0^{4\pi}\sin(x)\,dx\right) \\ &= \frac{\cos(\pi k/3)}{2}. \end{aligned}$$

(4 marks)

(e) Is $X(t)$ stationary?
Yes. (1 mark. Note though that if θ is taken as known, then $X(t)$ is purely deterministic and is not stationary.)

(f) Explain clearly how and why the spectral density function (also known as the spectrum) of $X(t)$ behaves at frequency $\pi/3$.
Recall the spectrum is defined as

$$f(\omega) = \frac{1}{\pi}\left(\gamma(0) + 2\sum_{k=1}^{\infty}\gamma(k)\cos(\omega k)\right).$$

For $X(t)$, *at* $\omega = \pi/3$,

$$f(\pi/3) = \frac{1}{\pi}\left(\frac{1}{2} + \sum_{k=1}^{\infty}\cos^2(\pi k/3)\right)$$

which does not converge since $\cos^2(\pi k/3) \nrightarrow 0$ *as* k *increases. Hence the spectrum is not defined at* $\pi/3$. (2 marks)

14. Let $X(t)$ and $Y(t)$ be stationary processes. Define the *cross covariance* function, $\gamma_{XY}(k)$, at lag k as

$$\gamma_{XY}(k) = \text{Cov}(X(t), Y(t+k)).$$

(a) In general, is $\gamma_{XY}(k) = \gamma_{XY}(-k)$? Explain your answer.
In general

$$\begin{aligned}\gamma_{XY}(-k) &= \text{Cov}(X(t), Y(t-k)) \\ &= \text{Cov}(Y(t-k), X(t)) \\ &= \text{Cov}(Y(s), X(s+k)) \\ &= \gamma_{YX}(k),\end{aligned}$$

and this does not necessarily equal $\gamma_{XY}(k)$. (2 marks, 1 for answer, 1 for explanation)

(b) Suppose that $X(t)$ and $Y(t)$ are independent stochastic processes. Is $\gamma_{XY}(k) = \gamma_{XY}(-k)$? Explain your answer.
In the case that $X(t)$ *and* $Y(t)$ *are independent,* $\gamma_{XY}(k) = 0$ *for all* k, *and so* $\gamma_{XY}(k) = \gamma_{XY}(-k)$ *does hold for all* k. (2 marks, 1 for answer)

(c) Suppose now that

$$\begin{aligned}X(t) &= Z_1(t) + 0.3Z_1(t-1) + 0.1Z_2(t-1) \\ Y(t) &= Z_2(t) + 0.7Z_1(t-1) + 0.5Z_2(t-1)\end{aligned}$$

where $Z_1(t)$ and $Z_2(t)$ are independent white noise processes, each with mean zero and variance 1. Find $\gamma_{XY}(k)$ for all k.

$$\begin{aligned}\gamma_{XY}(k) &= \operatorname{Cov}\left(Z_1(t) + 0.3Z_1(t-1) + 0.1Z_2(t-1),\right. \\ &\qquad \left. Z_2(t+k) + 0.7Z_1(t+k-1) + 0.5Z_2(t+k-1)\right) \\ &= \begin{cases} 0.3 \times 0.7 + 0.1 \times 0.5 = 0.26 & k = 0 \\ 0.7 & k = 1 \\ 0.2 & k = -1 \\ 0 & \text{else} \end{cases}\end{aligned}$$

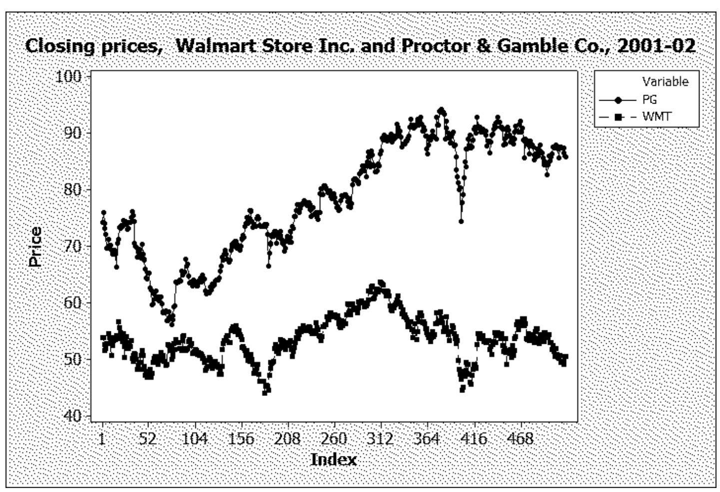
(4 marks)

(d) Find $\gamma_{YX}(k)$ for all k.

$$\gamma_{YX}(k) = \begin{cases} 0.26 & k = 0 \\ 0.7 & k = -1 \\ 0.2 & k = 1 \\ 0 & \text{else} \end{cases}$$

(2 marks)

(e) The plot below shows the closing prices of unit shares in Proctor and Gamble Co. ($X(t)$ say) and Walmart Store Inc. ($Y(t)$ say) over the years 2001 and 2002.



Which of the following do you think would be the most likely values (to 2 d.p.) for the *sample cross correlation function* $r_{XY}(k)$ for lags $k = -4, \ldots, +4$? Circle your choice.

i.

k	–4	–3	–2	–1	0	1	2	3	4
$r_{XY}(k)$	0.88	0.90	0.91	0.93	1.00	0.93	0.91	0.90	0.88

ii. *

k	–4	–3	–2	–1	0	1	2	3	4
$r_{XY}(k)$	0.47	0.47	0.47	0.47	0.47	0.46	0.45	0.44	0.44

*

iii.

k	–4	–3	–2	–1	0	1	2	3	4
$r_{XY}(k)$	–0.88	–0.90	–0.91	–0.93	0.95	0.93	0.91	0.90	0.88

iv.

k	–4	–3	–2	–1	0	1	2	3	4
$r_{XY}(k)$	0.88	0.90	0.91	0.93	0.98	–0.93	–0.91	–0.90	–0.88

v.

k	–4	–3	–2	–1	0	1	2	3	4
$r_{XY}(k)$	0.44	0.44	0.45	0.46	1.00	0.46	0.45	0.44	0.44

(2 marks)