1. (a) There are two definitions of stationarity.

   Strict stationarity: For any choice of $k > 0$, $t_1, \ldots, t_k$ and u, the joint distribution of $Y_{t_1}, \ldots, Y_{t_k}$ is the same as that joint distribution of $Y_{t_1+u}, \ldots, Y_{t_k+u}$.

   Weak Stationarity: The mean of the series is constant, the variance exists and the autocovariance between Y_t and Y_s is a function only of $t - s$.

   (b) A series is causal, if it has a moving-average representation of the form

$$Y_t = \sum_{u=0}^{\infty} \psi_u \varepsilon_{t-u}$$

   for some white-noise series ε_t. In simple terms, Y_t depends only on present and past outcomes, not future ones.

   (c) A series is an AR(1) series if it can be represented in the form

$$Y_t = \phi Y_{t-1} + \varepsilon_t$$

   where ε_t is a white noise series.

   (d) An AR(1) series will be stationary and causal if $|\phi| < 1$.

   (e) The standard trick says to write down the definition of Y_{t+u} multiply by Y_t and take expectations.

$$\begin{aligned} \gamma(u) &= E[Y_{t+u} X_t] \\ &= E[(\phi Y_{t+u-1} + \varepsilon_{t+u}) Y_t] \\ &= E[\phi Y_{t+u-1} Y_t] + E[\varepsilon_{t+u} Y_t] \\ &= \gamma(u-1), \end{aligned}$$

   for $u > 0$. To start this recursion, note that

$$var[Y_t] = \phi^2 var[Y_{t-1}] + var[\varepsilon_t],$$

   or

$$\gamma(0) = \phi^2 \gamma(0) + \sigma_\varepsilon^2,$$

   or

$$\gamma(0) = \frac{\sigma_\varepsilon^2}{1 - \phi^2}.$$

   Combining the results above gives

$$\gamma(u) = \frac{\sigma_\varepsilon^2 \phi^u}{1 - \phi^2} \qquad \text{for } u = 1, 2, \ldots$$

   The full result follows because $\gamma(u) = \gamma(-u)$.

   (f) Starting with

$$Y_t = \phi Y_{t-1} + \varepsilon_t$$

   we form differences

$$\begin{aligned} Y_t - Y_{t-1} &= (\phi Y_{t-1} + \varepsilon_t) - (\phi Y_{t-2} + \varepsilon_{t-1}) \\ &= \phi(Y_{t-1} - Y_{t-2}) + \varepsilon_t - \varepsilon_{t-1} \end{aligned}$$

If we let Z_t denote the diffenced series, it satisfies

$$Z_t = \phi Z_{t-1} + \varepsilon_t - \varepsilon_{t-1}$$

which is an ARMA(1,1) series.

(g) We can compute the variance of the differenced series as follows.

$$\begin{aligned} var[Y_t - Y_{t-1}] &= var[Y_t] + var[Y_{t-1}] - 2cov(Y_t, Y_{t-1}) \\ &= \gamma(0) + \gamma(0) - 2\gamma(1) \\ &= \frac{\sigma_\varepsilon^2(1 + 1 - 2\phi)}{1 - \phi^2} \\ &= \frac{2\sigma_\varepsilon^2(1 - \phi)}{(1 + \phi)(1 - \phi)} \\ &= \frac{2\sigma_\varepsilon^2}{1 + \phi} \end{aligned}$$

(h) The formula is

$$(1 - L)^d(1 - \phi_1 L^1 - \cdots - \phi_p L^p)Y_t = (1 + \theta_1 L^1 + \cdots + \theta_q L^q)\varepsilon_t$$

or

$$(1 - L)^d \phi(L) Y_t = \theta(L)\varepsilon_t.$$

2. (a) Differencing is a way of dealing with nonstationarity in time series. It can be used to deal with the presence of slowly varying trends or regular seasonal fluctation. Typically, the need for differencing is judged by examining a plot of the data for trends (either simple or seasonal) and perhaps examining the acf of the data looking for slow decay at either simple or seasonal lags.

In the case of the unemployment series, the slow (simple lag) variation looks to be the biggest effect. This should be removed first and then the resulting series checked for a remaining seasonal effect.

(b) There appears to be slow variation at simple lags in the acf with cutoff in the pacf after lag 3. There is what appears to be exponential decay at multiple of the seasonal period in the pacf and sharp cutoff after one seasonal lag in the pacf. This is a strong indicator that the appropriate model is ARIMA$(3, 1, 0) \times (0, 1, 1)_{12}$ (this is a very cool example).

(c) The model is:

$$\begin{aligned} &(1 - L)(1 - L^{12})(1 - \phi_1 L - \phi_2 L^2)(1 - \Phi_1 L - \Phi_2 L^2)Y_t \\ &\quad = (1 - \theta_1 L - \theta_2 L^2)(1 - \Theta_1 L - \Theta_2 L^2)\varepsilon_t \end{aligned}$$

(d) The `sar2` term is clearly not significant and can be dropped. Once this has been done the model should be refitted and the coefficients rexamined. Once all non-significant coefficients have been eliminated a check should be made to see if any other terms should be added to the model. Blind application of criteria such as AIC should not be used.

(e) The ACF for the residuals indicates whether there are any correlations which are significantly different from 0. The Ljung-Box p-values indicate whether there are groups of correlations in the residuals which are significant. The raw residual plot can indicate large outliers and other patterns which do not show up in the correlation plots.

(f) The two lower plots indicate that there are no significant correlations present in the residuals. The standardized residuals do show some evidence of lack of fit. In particular there is a large spike present in the mid seventies which is probably due to the oil shock(s) then. There is also an effect of the presidency of Ronald Reagan on unemployment. This probably don't effect the forecasts significantly, but we should keep in mind that particular events can produce significant deviations from the forecasts.

(g) There is evidence of a small on-going seasonal pattern with two peaks each year in the series. One in January (post Christmas) and another in June/July (post graduation?). However, the confidence intervals around the forecasts indicate that there is a great deal of uncertainly present in the series and that forecasts out beyond three or four months are not very reliable.

3. (a) i. The impulse response is

$$a(u) = \begin{cases} 1/2 & u = -1 \\ -1/2 & u = 1 \\ 0 & \text{otherwise.} \end{cases}$$

ii. The transfer function is

$$\begin{aligned} A(\lambda) &= \sum_{u=-\infty}^{\infty} a(u) e^{-i\lambda u} \\ &= e^{i\lambda}/2 - e^{-i\lambda}/2 \\ &= (e^{i\lambda} - e^{-i\lambda})/2 \\ &= (2i \sin\lambda)/2 \\ &= i \sin\lambda. \end{aligned}$$

iii. A graph of this transfer function $A(\lambda)$shows that it is zero at $\lambda = 0$ and $\lambda = \pi$ and has a peak at $\lambda = \pi/2$. This means that the filter will tend to eliminate very high and very low frequencies and to pass through cosines at frequencies at about $\pi/2$ (i.e. 4 cycles per unit time).

(b) i. The power spectrum is defined to be

$$f_{XX}(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \gamma(u) e^{-i\lambda u}.$$

ii. The relationship is

$$f_{YY}(\lambda) = |A(\lambda)|^2 f_{XX}(\lambda)$$

where $A(\lambda)$ is the transfer function of the filter.

iii. The periodogram is asymptotically distributed as $f_{XX}(\lambda)\chi_2^2/2$ or, equivalently, exponential with mean $f_{XX}(\lambda)$. If $\lambda \neq \mu \mod 2\pi$, then $I_{XX}^T(\lambda)$ and $I_{XX}^T(\mu)$ are asymptotically independent.

iv. The periodogram is an inconsistent estimator of the power spectrum. On its own it provides a very spikey estimate because of the asymptotic independence of its values. The estimate can be improved by smoothing the periodogram with a moving average. This can be set up in such a way that it produces a consistent estimator.

(c) i. The model to be fitted is

$$Y(t) = \sum_{u=-\infty}^{\infty} a(u)X(t-u) + \varepsilon(t).$$

where $\{a(u)\}$ is the impulse response of the filter and $\varepsilon(t)$ is a "noise" series (not necessarily white-noise) which is independent of $X(t)$.

ii. The coherence between $X(t)$ and $Y(t)$ is defined by

$$R_{YX}^2(\lambda) = \frac{|f_{YX}(\lambda)|^2}{f_{YY}(\lambda)f_{XX}(\lambda)}.$$

The coherence provides a measure of the degree of association between the two series at frequency λ. It satisfies

$$0 \leq R_{YX}^2(\lambda) \leq 1.$$

4. (a) The first two plots show the power spectra of the input and output series. The spectra show a great deal of similarity at low frequencies, but the output series shows a higher level of high frequency noise.

The second and third plots show the gain and phase of the fitted filter. The filter passes through low frequencies and stops higher frequencies — in other words it is a smoothing filter. The phase shows a negative linear trend at low frequencies which show that very low frequencies are being delayed relative to moderate low frequencies.

The fifth and sixth plots show the choherence and residual spectrum. The coherence shows a high degree of relations ships for frequencies out to about 0.1. The coherence then drops, but shows evidence of significance across the full range of frequencies. The residual spectrum does not seem to be white noise, but is whiter than the input and output series.

The last plot shows the fitted filter coefficients. The plot indicates that the output is obtained from the input series by averaging the current and prior 8 or 9 values of the input series, with the weights appearing to decrease in a linear fashion. This is, indeed, a smoothing filter with properties in accord with those described above.

(b) The first plot shows a basic cosinusoid wave whose amplitude is being modulated by another, lower-frequency, cosinusoid. In other words it has the form

$$Y(t) = A(t) \cos \lambda_c t$$

where $A(t) = \cos \lambda_m t$ with $\lambda_m << \lambda_c$.

The second series appears to show two significant peaks at close frequencies of about 0.3, along with a number of possible harmonics. The power in these two components is roughly equal, which means that the associated cosinusoids have roughly equal amplitudes. This suggests a model of the form

$$Y(t) = A(\cos \lambda_1 t + \cos \lambda_2 t)$$

with $\lambda_1 \approx \lambda_2$.

The models can be seen to be very similar (perhaps identical) when we recognise that

$$2 \cos \theta \cos \phi = \cos(\theta - \phi) + \cos(\theta + \phi).$$

This suggests that $\lambda_c = \lambda_2 - \lambda_1$ and $\lambda_m = \lambda_1 + \lambda_2$.