Deduct marks for each major thing missed, and for each wrong statement. In general, be relatively generous if the answer makes sense and contains the main ideas.

1. *The AIC is better than the MSE for selecting a forecasting model.*

    - The MSE on the training set only measures goodness of fit, not forecast accuracy. $\boxed{1}$
    - The AIC is also a measure of goodness of fit on the training set, but includes a penalty for the number of parameters in the model. $\boxed{1}$
    - The MSE on a test set is a measure of forecast accuracy, but the test set may be too small to be reliable. The MSE on a cross-validated set is better but is computationally expensive. $\boxed{1}$
    - The AIC is asymptotically equivalent to one-step time series cross-validation. $\boxed{1}$
    - The AIC can only be compared within the same model class, whereas MSE on a test set, or on a cross-validated set, can be used to compare different model classes. $\boxed{1}$

2. *The Ljung-Box test is useful for selecting a good forecasting model.*

    - The Ljung-Box test is used to test whether the residuals from a model are white noise. $\boxed{1}$
    - It should not be used to select a forecasting model. $\boxed{1}$
    - If the residuals are not white noise, then the model is not capturing all the information in the data, and can potentially be improved. $\boxed{1}$
    - A model that has been over-fitted may pass a Ljung-Box test, but be a poor forecasting model. $\boxed{1}$
    - It may not be possible to find a model that passes the Ljung-Box test, especially with a long time series. But the forecasts from a model that fails the Ljung-Box test may still be good. $\boxed{1}$

The exam contains FIVE sections. ALL sections must be completed. The exam is worth 100 marks in total.

Below are the State Space equations for each of the models in the ETS framework.

ADDITIVE ERROR MODELS

Trend	**Seasonal** **N**	 **A**	 **M**
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
A$_\mathbf{d}$	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$

MULTIPLICATIVE ERROR MODELS

Trend	**Seasonal** **N**	 **A**	 **M**
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A$_\mathbf{d}$	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

SECTION B

Figures 1–3 relate to the daily use of public transport in Canberra, from July 2019 – March 2024. The variable plotted is the total number of passenger boardings each day on all forms of public transport except for school buses.

1. *Using Figures 1–2, describe the daily passenger boardings for public transport in Canberra. Canberra had two major periods of "COVID-19 lockdowns" where there was substantially reduced travel, and has four school terms per year. Carefully comment on the interesting features of both plots, and how the lockdowns, school terms and other holidays are evident.*

    - There is strong weekly seasonality with lower boarding numbers on weekends, seen in Figure 1 but more clearly in Figure 2. 1
    - Some Mondays (Fig 2) have lower boarding numbers, probably associated with long weekends. 1
    - Fig 1 shows there is also some annual seasonality, with low boarding numbers at the end of the year and the beginning of the next year (summer holidays). 1
    - The school holidays (incl Easter) have a noticeable effect with lower boarding numbers during the holidays/terms breaks. 1
    - The two COVID-19 lockdowns are visible in the first half of 2020, and the second half of 2021. 1

3. You have been asked to provide forecasts for the next four weeks for daily passenger boardings. Consider applying each of the methods and models below to the data from March 2022 onwards. Comment, in a few words each, on whether each one is appropriate for forecasting the next two weeks of data. No marks will be given for simply guessing whether a method or a model is appropriate without justifying your choice.

10 marks

Start your response by stating: **suitable** or **not suitable**.

(a) Seasonal naïve method using weekly seasonality.
(b) Naïve method.
(c) An STL decomposition on the log transformed data combined with an ARIMA to forecast the seasonally adjusted component, and seasonal naïve methods for both seasonal components.
(d) Holt-Winters method with damped trend and multiplicative weekly seasonality.
(e) ETS(A,N,A).
(f) ETS(M,A,M) with annual seasonality.
(g) ARIMA(2,4,2)(1,1,0)$_7$ applied to the log transformed data.
(h) ARIMA(1,0,1)(1,1,0)$_7$ applied to the log transformed data.
(i) Regression with time and Fourier terms for both weekly and annual seasonality.
(j) Dynamic regression on the log transformed data with Fourier terms for the annual seasonality and a seasonal ARIMA model to handle the weekly seasonality and other dynamics.

Total: 20 marks

(d) *Holt-Winters method with damped trend and multiplicative weekly seasonality.*

- Might be suitable, but will miss annual seasonality and holidays. $\boxed{1}$

(e) *ETS(A,N,A).*

- Unsuitable. Seasonality is not additive. Also misses annual seasonality and holidays. $\boxed{1}$

(f) *ETS(M,A,M) with annual seasonality.*

- Unsuitable. ETS won't handle annual seasonality, and model doesn't capture weekly seasonality. $\boxed{1}$

SECTION C

1. *An ETS model is fitted to the time series, using only data from March 2022. Write down the equations for the model, including specifying the values of all model parameters.* $\boxed{3}$

$$
\begin{aligned}
y_t &= \ell_{t-1} s_{t-m}(1+\varepsilon_t) \\
\ell_t &= \ell_{t-1}(1+\alpha\varepsilon_t) \\
s_t &= s_{t-m}(1+\gamma\varepsilon_t) \qquad\qquad \varepsilon_t \sim N(0,\sigma^2)
\end{aligned}
$$

   where $\alpha = 0.208$, $\gamma = 0.0001$, $\sigma^2 = 0.0338$. $\boxed{2}$

2. *What features of the data has the model ignored?*
   - annual seasonality $\boxed{1}$
   - holidays $\boxed{1}$

3. *What does the value of γ tell you?*
   - seasonal pattern hardly changing over time. $\boxed{1}$

4. *The data, remainder (i.e., residuals), and estimated states, are shown below for the last week of observations, along with the forecasts for the next day. Show how the forecast mean and variance have been obtained, and give a 95% prediction interval for this day.*
   - Forecast mean: $\hat{y}_{T+1|T} = \ell_T s_{T+1-m} = 48.1 \times 1.17 = 56$ $\boxed{3}$
   - Forecast variance is 106. So a 95% prediction interval is $\boxed{2}$

$$56 \pm 1.96 \times \sqrt{106} = (36, 76)$$

5. *Some plots of the residuals are shown in Figure 4. Discuss what these tell you about the model?*
   - Significant serial correlation remaining in residuals. $\boxed{1}$
   - Lots of outliers, probably associated with holidays. $\boxed{1}$
   - Residuals are not normally distributed. $\boxed{1}$
   - The model has not captured the holiday effects, or the time series dynamics in the data. $\boxed{1}$
   - The model will give prediction intervals with incorrect coverage. $\boxed{1}$

6. *If you conducted a Ljung-Box test of the residuals using 14 lags, what do you think the p-value would be? Why?*
   - Close to zero $\boxed{1}$
   - Due to significant serial correlation. $\boxed{1}$