Forecasting Exam 2025: solutions

The exam contains FIVE questions. ALL questions must be answered. The exam is worth 100 marks in total.

SECTION A

*Respond to any **four** of the following topics. For each, demonstrate your understanding by presenting at least five key insights or explanations.*

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. *An STL decomposition into trend, seasonal and remainder components is only useful for forecasting when there are no cycles in the data.*

- In STL decomposition, the trend component may also capture cycles, especially if they are long-term. Hence, the statement is not true. 1
- STL allows both the trend/cycle and seasonal components to change over time through the `window` arguments in the `trend()` and `season()` components. This offers flexibility in capturing changes in these patterns. 1
- After decomposition, the resulting seasonally-adjusted time series and the seasonal components are forecasted separately. 1
- The seasonal patterns occur at regular intervals and are easy to forecast. 1
- In contrast, cycles contained in the seasonally-adjusted series, do not have a fixed period and can vary in length and amplitude. Hence, these are much harder to forecast. 1

2. *Forecasting assumes that the patterns in the past will continue in the future.*

- Forecasting models rely on the assumption that historical patterns such as trends and seasonality will continue into the future. However, they do not assume things will stay the same. 1
- Models like ETS use components estimated from past data to generate forecasts. These patterns are allowed to change over time. 1
- Models like ARIMA explore dynamics in the historical data and project these in the future to generate forecasts. They can model both levels but also differences. Models in differences require changes to continue in the future. 1
- Forecasts are conditional on past observed behaviour, which is observed in the training data. Their accuracy depends heavily on structural changes not occurring during the period we are trying to forecast. 1
- Prediction intervals are important as these account for and project uncertainty, recognising that future patterns will not exactly match the past. 1

3. *Benchmark forecasting methods are too simplistic to be used in practice.*

- Benchmark methods like the naïve and seasonal naïve are intentionally simple but they are effective in many situations. 1
- They provide a baseline to compare more complex forecasting methods and models and help us identify when added complexity may be unjustified. 1
- In some cases, especially with short-term or highly seasonal data, benchmark methods outperform more sophisticated techniques. 1
- These methods require minimal assumptions, making them useful in practice when data is limited. 1
- While they may not capture all nuances, dismissing them as too simplistic ignores their value in forecast accuracy evaluation. 1

4. *Box-Cox transformations are useful for making a time series stationary.*

- Box-Cox transformations are primarily used to stabilise the variance in a time series, which is a key requirement for stationarity. 1
- Heteroscedasticity can distort model estimates, and transforming the data helps improve model performance. 1
- The transformation includes common options like the log or square root, which are special cases of the Box-Cox transformation. 1
- While the Box-Cox transformation can help with variance stationarity, it does not address non-stationarity in the mean, which often requires differencing. 1
- Box-Cox transformations are often used in combination with differencing to achieve full stationarity before fitting ARIMA models. 1

5. *Time series cross-validation is not really necessary because we can just use the AICc.*

- AICc evaluates fit over the training set, penalised for model complexity, but does not directly assess out-of-sample forecast accuracy. 1
- Minimising the AIC assuming Gaussian errors is asymptotically equivalent to minimising one-step time series cross validation MSE. 1
- Time series cross-validation focuses on evaluating out-of-sample forecast accuracy, and it also allows for evaluating multi-step forecasts. 1
- Time series cross-validation can be very time-consuming and expensive, therefore possibly not feasible in practice. 1
- In cases where data is limited/short, AICc may be the only feasible option. However, AICc cannot be used across different classes of models. 1

6. *Never produce forecasts using a model that has correlated residuals.*

- Correlated residuals indicate that the model has not captured all the dynamics in the data. 1
- Accuracy of forecasts and particularly prediction intervals will be adversely affected resulting in incorrect coverage, especially if the correlations are strong. 1
- We aim to estimate models with uncorrelated residuals. However, this may not always be possible. In the case that the correlations are low, forecasts will be reasonable. 1
- Residual autocorrelation can be detected using diagnostic tools like the Ljung-Box test or ACF plots. 1
- It may be possible to revise a model to reduce autocorrelation in the residuals. But we should always aim to avoid overfitting the data. 1

Total: 20 marks

SECTION B

1. *Describe Figures 1–3. Carefully comment on the interesting features of the plots.*

Figure 1

- Time plot shows annual seasonality with electricity demand higher during the summer months due to air conditioning 1
- Winter months seem to also have higher demand compared to spring and autumn, due to heating. 1
- Noticeable very high spikes in demand are observed, particularly during summer months, which may represent outliers due to extreme events such as extremely hot and/or humid days. 1
- There appears to be some heteroscedasticity, with the variability in demand changing over time. For example, the demand fluctuations in early 2024 and 2025 are more pronounced than in previous years. 1

Figure 2

- The daily season plot shows strong daily seasonality. Demand peaks around 6pm when people return home after work. There is a first spike around 7am when people wake up in the morning. Demand is at its lowest between 10am and 2pm when people are at work. 2

Figure 3

The daily seasonal pattern from Figure 2 recurs across each day of the week in Figure 3. Weekdays show similar shapes with pronounced morning and evening peaks, while weekends have slightly lower demand overall. The morning bump around 7am, likely reflecting workday routines, seems to be absent on weekends.

1

2. *The data has been aggregated to daily. For the STL decomposition in Figure 4, briefly discuss what is shown in each panel. How has the annual seasonality been handled?*

The STL decomposition shows the observed daily electricity demand in the top panel. The second panel shows the smooth trend/cycle component. The third panel displays a weekly seasonal pattern, repeating every 7 days. The bottom panel shows the remainder with a few potential outliers. 2

Since the decomposition uses weekly seasonality, annual seasonality is not captured and remains in the trend component. 1

3. *You have been asked to provide forecasts for the next two weeks using the daily data. Consider applying each of the methods and models below. 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.*

Be flexible in marking and reward appropriate remarks.

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

(a) *Seasonal naïve method using weekly seasonality.*
   - Suitable. It will capture the weekly seasonality but will miss the annual seasonality. Reasonable for short-term forecasting. 1

(b) *Naïve method.*
   - Not suitable. It will not even capture weekly seasonality. 1

(c) *The STL decomposition shown in Figure 4, combined with an ARIMA to forecast the seasonally adjusted component, and seasonal naïve method for the seasonal component.*
   - Suitable. The ARIMA will capture and project the trend/cycle and reminder in the seasonally adjusted component, and the seasonal naïve will reflect the weekly seasonality which will be fine for short-term forecasting. 1

(d) *Holt-Winters method with no trend and additive weekly seasonality.*
   - Suitable. Will miss annual seasonality but ok for short-term forecasting. 1

(e) *ETS(A,N,A) with annual seasonality.*
   - Not suitable. Not possible to fit an ETS with seasonal period 365. Will miss weekly seasonality 1

(f) *ETS(M,A,M) with weekly seasonality.*
   - Suitable. Multiplicative weekly seasonality and adaptable additive trend seems very reasonable for short-term forecasting. 1

(g) *ARIMA(2,0,1)(1,0,0)*$_{52}$.
   - Not suitable. The data is daily not weekly. 1

(h) *ARIMA(1,1,2)(2,0,0)*$_{7}$.
   - Suitable. The seasonal AR(2) will pick up the weekly seasonality. OK for short-term forecasting but misses annual seasonality. 1

(i) *Dynamic regression with Fourier terms for the annual seasonality and a seasonal ARIMA model to handle the weekly seasonality and other dynamics.*
   - Suitable. The model will capture both annual (through Fourier terms) and weekly (through ARIMA) seasonality. 1

(j) *Dynamic regression with Fourier terms for the weekly seasonality and a seasonal ARIMA model to handle the annual seasonality and other dynamics.*
   - Not Suitable. An ARIMA model is not suitable for capturing annual seasonality in daily data $(P,D,Q)_{365}$. 1

Total: 20 marks

SECTION C

1. *Write down the equations for the fitted model, and explain how the output above relates to the model parameters.*

The fitted model is an `ETS(M,Ad,M)` model, with multiplicative errors, additive damped trend, and multiplicative seasonality.

$$\begin{aligned}
y_t &= (\ell_{t-1} + \phi b_{t-1}) s_{t-m}(1 + \varepsilon_t) \\
\ell_t &= (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t) \\
b_t &= \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t \\
s_t &= s_{t-m}(1 + \gamma\varepsilon_t)
\end{aligned}$$

$\boxed{2}$

where $m = 7$ (weekly seasonality), $\varepsilon_t \sim N(0, \sigma^2)$ and $\hat{\sigma}^2 = 0.001$. $\boxed{1}$

The estimated smoothing parameters are:

- $\alpha = 0.9792$
- $\beta = 0.0006$
- $\gamma = 0.0002$
- $\phi = 0.9711$ $\boxed{2}$

The initial states are estimated to be:

- $\ell_0 = 1870$
- $b_0 = 3.330$
- $s_0, \dots, s_{-6}$: 1.011, 1.022, 1.026, 1.027, 1.015, 0.945, 0.955 $\boxed{1}$

2. *Plots of the residuals are shown in Figure 4. Discuss what these tell you about the model.*

- Time plot seems to display random residuals - there are however some large residuals which possibly may be outliers relating to extreme weather events that are not accounted for by the model. $\boxed{1}$
- ACF plot shows some significant autocorrelations, lags 2, 3 and 4 but also on the seasonal lags, 14 and 21, suggesting there are some dynamics not captured. $\boxed{1}$
- Histogram shows that that residuals may be approximately normally distributed. $\boxed{1}$
- The significant residual autocorrelations and outliers will significantly affect the reliability and the width of the prediction intervals. $\boxed{1}$

3. *If a Ljung-Box test was conducted on the residuals from this model, what would you expect the p-value to be? Justify your answer.*

I would expect the p-value from the Ljung-Box test to be low, less than 0.05, indicating that the residuals are not white noise. $\boxed{1}$

This is because the ACF plot of the residuals shows significant autocorrelation at multiple lags (especially at lags 2, 3, 4, and 14). The Ljung-Box test checks for overall/joint autocorrelation, and the presence of these significant autocorrelations suggests that the test would reject the white noise hypothesis. $\boxed{1}$

4. *The last few values of the model states are shown below. Use these, along with the model output shown earlier, to compute the point forecast for the next day.*

$$y_{T+1|T} = (\ell_T + \phi b_T)s_{T-6}(1 + e_{T+1})$$

$\boxed{1}$

$$\hat{y}_{T+1|T} = (\ell_T + \phi b_T)s_{T-6}(1 + e_{T+1})$$

$\boxed{1}$

$$\hat{y}_{T+1|T} = (1930 + 0.971 \times 0.00577) \times 1.03 = 1981$$

$\boxed{4}$

Note: rounding causes some inaccuracy. They will obtain 1988 if they use the printed figures. Mark that as correct.

5. *Forecasts and prediction intervals for the next two weeks are shown in Figure 6. How many of the 80% intervals do you expect to contain the actual demand values?*

If the coverage is correct, you expect 80%$= 0.8 \times 14 = 11.2$ of the intervals on average to contain the future actual observed values of demand. $\boxed{2}$

Total: 20 marks

SECTION D

1. *Write down the equations for the model using backshift notation, including specifying the values for all model parameters.*

   The fitted model is ARIMA(1,1,2)(2,0,0)[7]. Using backshift notation, the model is:

   $$(1-\phi_1 B)(1-\Phi_1 B^7 - \Phi_2 B^{14})(1-B)y_t = (1+\theta_1 B + \theta_2 B^2)\varepsilon_t$$

   where $\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$. $\boxed{3}$

   Substituting in the estimated coefficients:

   $(1-0.5458B)(1-0.3508B^7 - 0.2071B^{14})(1-B)y_t = (1-0.6655B-0.2877B^2)\varepsilon_t$ and $\hat{\sigma}^2 = 5442$ $\boxed{2}$

2. *Explain how the model above has been selected.*

   The model has been automatically selected using the `ARIMA()` function from the `fable` package. After establishing stationarity the function searches over possible combinations of non-seasonal and seasonal ARIMA terms and selects the model that minimises the corrected Akaike Information Criterion (AICc). This ensures a good balance between model fit and complexity, i.e., the number of parameters estimated. $\boxed{2}$

3. *Using the plots in Figure 7 choose an alternative model? Justify your choices.*

   - The time series has been differenced at lag 7 to remove weekly seasonality. Hence $D = 1$. $\boxed{1}$
   - Focusing on the seasonal lags, the ACF shows a significant spike at lag 7 while the PACF decays, which is characteristic of a seasonal MA(1) process. $\boxed{2}$
   - For the non-seasonal lags, the PACF shows two significant spikes followed by a gradual decay, while the ACF is also slowly decaying. This suggests a non-seasonal AR(2). $\boxed{2}$
   - Hence, a well-justified alternative model would be an ARIMA(2,0,0)(0,1,1)[7]. $\boxed{1}$

   Allow for other reasonable models if they are well-justified.

4. *Comment on the residuals shown in Figure 8. What do you conclude in terms of the fit of the model. Would forecasts generated from the model be reliable based on the residuals?*

- The time plot of the residuals appears mostly random, indicating the model captures much of the structure in the data. However, there are some large residuals, particularly during the summer of 2025, which may be outliers due to extremely hot and/or humid days. $\boxed{1}$
- The ACF shows some significant spikes at long lags. The ones that would lead me to adjust and possibly improve the model are the significant spikes at the seasonal lags of 14 and 28. This indicates that the current model's seasonal component may be mis-specified or insufficient, and could be improved by adjusting the seasonal AR or MA terms. $\boxed{1}$
- The histogram shows pretty well-behaved residuals and approximately normal. $\boxed{1}$
- Overall, the model is likely to provide reasonable short-run point forecasts, but the presence of residual autocorrelation and outliers means that prediction intervals may not have correct coverage. Therefore, it would be advisable to revise the seasonal structure to improve forecast reliability. $\boxed{2}$

5. *The point forecast for the next day is 1982 MW. Compute the corresponding 95% prediction interval.*

   $\hat{y}_{T+1|T} \pm 1.96\hat{\sigma}$, where $\hat{\sigma} = \sqrt{5442} \approx 73.77$. Hence, $\boxed{1}$

   $$1982 \pm (1.96 \times 73.77) = 1982 \pm 144.59$$

   Thus, the 95% prediction interval is approximately: $(1837,\ 2127)$ MW. $\boxed{1}$

Total: 20 marks

SECTION E

The time series shown in Figure 4 is split into a training set and a test set, where the test set comprises the last 2 weeks of available data. A dynamic regression model is fitted to training data, along with an ETS and an ARIMA model.

1. *The model details for the dynamic regression are shown below. Write down the equations for the model. There is no need to give numerical values for the coefficients.*

   The model is a dynamic regression with Fourier terms to capture weekly seasonality, and ARIMA(1,1,2) errors:

   $$y_t = \sum_{k=1}^{3}\left[\beta_k \sin\left(\frac{2\pi kt}{7}\right) + \gamma_k \cos\left(\frac{2\pi kt}{7}\right)\right] + \eta_t,$$

   where η_t follows an ARIMA(1,1,2) model: 2

   $$(1-\phi_1 B)(1-B)\eta_t = (1+\theta_1 B + \theta_2 B^2)\varepsilon_t,$$

   with $\varepsilon_t \sim \text{N}(0,\sigma^2)$. 2

2. *Which coefficients in the dynamic regression model relate to weekly seasonality? What do the remaining coefficients handle?*

   - The coefficients that relate to weekly seasonality are those associated with the Fourier terms: β_k, γ_k for $k = 1,\dots,3$. 2
   - The remaining coefficients, $\phi_1, \theta_1, \theta_2$, relate to the ARIMA(1,1,2) error structure. These terms account for short-term autocorrelations and the non-stationary dynamics in the residuals that are not explained by the regression component. 1

3. *We use all three models to forecast the next 2 weeks of data, and compare the results against the actual values. Which model produces the most accurate forecasts for the test period? Justify your answer using the forecast accuracy metrics. Then, compare this result with the residual diagnostics presented in Sections C and D. How do you explain the difference between forecast accuracy and model fit?*

   - The ETS model produces the most accurate forecasts for the 2-week test period, as shown by the lowest RMSE, MAPE, MASE and RMSSE values. All accuracy measures unanimously point to this. 1
   - Comparing the residual diagnostics, the ARIMA model had better-behaved residuals than the ETS model. The ETS residuals displayed significant autocorrelations, suggesting it failed to capture all dynamics in the training data. This mainly affects forecast uncertainty and the reliability of prediction intervals. 1

   One mark for either one of the following 1

   - This highlights the important distinction between in-sample fit and out-of-sample forecast performance. A model like ARIMA may fit the training data better, yet still yield less accurate short-term forecasts than a more parsimonious model like ETS.
   - The differences may arise from model flexibility, overfitting, or the ability of the simpler ETS structure to generalise better over a short forecast horizon.

4. *How would you extend each of the three models to also handle the annual seasonality?*

   A maximum of <u>4 marks</u> comprised of the following points:

   - Add <u>Fourier terms for annual seasonality</u> to the regression: 2
   -

   $$\sum_{k=1}^{K}\left[\alpha_k \sin\left(\frac{2\pi kt}{365}\right) + \delta_k \cos\left(\frac{2\pi kt}{365}\right)\right]$$

   1

- where K controls the number of harmonics used to approximate the yearly seasonal pattern. 1
- Fourier terms approximate the seasonality such that the seasonality does not change over time 1
- Use a STL decomposition (+1) and separately model the annual seasonal pattern (with dummy seasonality (+0.5), or ideally fourier terms (+1)). 2
- Mention how adding annual seasonality to the ETS model could prove difficult 1

5. **ETC3550 only**

Briefly describe each of the accuracy measures shown in the table above, and what they are measuring.

- MAE (Mean Absolute Error) is the average of the absolute forecast errors, and RMSE (Root Mean Squared Error) is the square root of the average squared forecast errors. While MAE treats all errors equally, RMSE penalises larger errors more heavily, making it more sensitive to outliers. MAE and RMSE are scale-dependent, and thus not suitable for comparing accuracy across time series with different units or magnitudes. 2

- MAPE (Mean Absolute Percentage Error) is the average of the absolute percentage errors. It expresses forecast accuracy as a percentage, making it widely used and easily interpretable. MAPEs are undefined when actual values are zero and are highly sensitive to values close to zero due to their percentage-based formulation. 2

- MASE (Mean Absolute Scaled Error) and RMSSE (Root Mean Squared Scaled Error) use absolute and squared errors, respectively, scaled relative to the MAE of an in-sample naïve forecast. Values below 1 indicate that the model outperforms the naïve benchmark. MASE and RMSSE are scale-independent and robust, making them preferable when comparing performance across different series. 2

6. **ETC5550, ETF3231, and ETF5231 only**

Energy experts tell you that electricity demand in Queensland increases on hot and humid days due to air-conditioning. Explain how you could use this information to improve your model. What complications would this add to the forecasting process?

Improve the model by incorporating *temperature* and *humidity* as predictor variables. Combined with the Fourier terms and ARIMA error process, the dynamic regression model will also include:

$$\beta_1 \text{Temperature}_t + \beta_2 \text{Humidity}_t$$

Including these variables allows the model to capture variation in electricity demand driven by weather conditions, improving forecast accuracy, particularly during hot and humid days. 2

A key complication is that forecasting demand now requires reliable forecasts of temperature and humidity. If these weather forecasts are inaccurate, the resulting demand forecasts will also be affected. This adds an extra layer of uncertainty. For short-term forecasting (e.g., 1–2 weeks ahead), this is should be manageable. 1 1 1

We should note, that the relationship between temperature, humidity, and demand may not be linear. For example, demand may increase sharply only beyond certain temperature thresholds. This would require incorporating nonlinear terms or interaction effects into the model. 1

Total: 24 marks