Forecasting Exam 2023: solutions

SECTION A

*Write about a quarter of a page each on any **four** of the following topics. Clearly state if you agree or disagree with each statement.*

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. *Linear regression models are simplistic because the real world is nonlinear.*

- This statement is true, but that does not make linear models not useful. $\boxed{1}$
- A linear regression model is often a useful although simple approximation to reality that can work well. $\boxed{1}$
- If there is not enough data to estimate the nonlinearity, particularly if the nonlinearity is too mild to estimate well, then a linear model is often the best approach. $\boxed{2}$
- The linearity is in the parameters, not in the functional form. So nonlinear relationships can be modelled using a linear model. $\boxed{1}$

2. *MAE, RMSE, MAPE and MASE are all similar measures; hence it does not matter which one we use.*

- This statement is not true, as these are very different measures with different properties. $\boxed{1}$
- The first two measures are scale dependent and if the series are measured on different scales these cannot be used for comparison across these series. The latter two measures are scale independent $\boxed{1}$
- With MAPE we need to be careful as it is not specified for series with zeros and it will be very skewed if we have observations close to zero. $\boxed{1}$
- MASE is robust to this as the scaling factor comes from the MAE over the training set of a simple method such as the naive for non-seasonal time series. $\boxed{2}$

3. *Always choose the model with the smallest sum of squared errors.*

- This statement is not true, as simply considering sum of squared errors is problematic, especially for the purpose of forecasting. $\boxed{1}$
- Selecting a model based on SSE is problematic because models can be built to over-fit the data which has an adverse effect on forecasting. $\boxed{1}$
- For example, in a regression model adding another variable will always reduce the sum of squared errors, even if the additional variable is of no value in forecasting. We can always make an ARIMA model more complicated to fit the data better. $\boxed{1}$
- Instead, it is best to choose models based on information criteria such as the AICc. Such criteria penalise the sum of squared errors to take into account the number of parameters in the model / to take into account the complexity of the model. $\boxed{2}$

4. *A good test of whether a model will produce good forecasts is that the residuals are stationary.*

- This is a false statement as we need more than stationary residuals. $\boxed{1}$
- Residuals also have to be white noise (i.e., uncorrelated) so that all the serial correlation/signal/information in the data has been captured. $\boxed{1}$
- If the residuals contain autocorrelation, then the prediction intervals will be incorrect, and the forecasts will not be fully efficient. $\boxed{2}$
- Furthermore, we need to make sure that that model is as parsimonious as possible, hence it is not over-fitting the data. $\boxed{1}$

5. *Forecasts should never give the same value for all forecast horizons.*

- This statement is false, as it is possible to have point forecasts of the same value. $\boxed{1}$
- Point forecasts are usually the means of the future forecast distributions. $\boxed{1}$
- Hence, it is possible for the means to be constant for all horizons for various models $\boxed{1}$
- For example, this happens with a white noise or a random walk model, an ETS(A,N,N) or an ETS(M,N,N) model, an MA(1) or an ARIMA(0,1,1) model. $\boxed{2}$

6. *Regression models are not useful for forecasting because we always need to provide forecasts of the predictors.*

- This statement is false. $\boxed{1}$
- We can specify different types of models: for example ones with only lags of the predictors; or ones with deterministic predictors such as trend and/or seasonal dummies or other dummies, or Fourier terms that do not need forecasts of predictors. $\boxed{2}$
- When the predictors are unknown in the future, you do have to forecast them, or some may be fully controlled by the policy maker (e.g., how much the company will spend on advertising next month) or generate alternative plausible scenarios. Scenario based forecasting is very useful for decision making. $\boxed{2}$

[Total: 20 marks]

— END OF SECTION A —

SECTION B

1. *Using Figures 1–3, describe the weekly mortality rate for Australia. Carefully comment on the interesting features of all three plots. [6 marks]*

- Time plot shows annual seasonality (with peaks during the winter weeks) $\boxed{1}$
- Time plot shows an increasing trend over the last couple of years - covid effect $\boxed{1}$
- Time plot shows a higher peak compared to other pre-Covid years for 2017; a different shape (fewer deaths over the winter period) for the winter peak of 2020; and an unusual summer peak in 2022. $\boxed{1}$
- Season plot verifies the annual seasonality, with peaks during the weeks of June to September, the increasing trend with 2021 and 2022 being higher than the other years, the unusually high peak for 2017 compared to the other pre-Covid years, the lower rate for the winter weeks for 2020; and the unusual summer peak of 2022. $\boxed{1}$
- STL decomposition verifies the features we have already seen in the trend panel, both in terms of an unusual bump in trend in 2017, the lower trend in the winter of 2020 and the increasing trend in the last few years $\boxed{1}$
- The remainder shows a few outliers with the most pronounced at the beginning of 2022 $\boxed{1}$

2. *For the STL decomposition shown in Figure 3, discuss the effect of the window sizes chosen for the trend and seasonal components. How would the results have changed with smaller or larger values chosen in each case? [4 marks]*

- The trend window includes 9 weeks. This is quite small. As a result the trend is not very smooth and possibly contains outliers such as the 2022 bump. Increasing this would make the trend smoother and possibly push to the remainder some possible outliers. $\boxed{2}$

- The seasonal window is set to 'periodic' and as a result the seasonal component is not allowed to change over time. A smaller window would allow the seasonal component to vary but we only have 8 years of data so the window cannot vary by much . $\boxed{2}$

3.(GA: I think we need to be lenient here if they pick up/comment on short v long run forecasts. We should keep this in mind)

(a) *Seasonal naïve method using annual seasonality.*

- Suitable. It would generate the annual seasonality. Not ok if the trend continues. $\boxed{1}$

(b) *Seasonal naïve method using weekly seasonality.*

- Not suitable. There is no weekly seasonality. $\boxed{1}$

(c) *An STL decomposition with an ETS to forecast the seasonally adjusted component, and seasonal naïve for the seasonal component.*

- Suitable. This would work well as the possible trend will also be projected. $\boxed{1}$

(d) *Holt's method with damped trend.*

- Not suitable. It would not allow for the annual seasonality. $\boxed{1}$

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

- Not suitable. It would not allow for the annual seasonality. $\boxed{1}$

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

- Not suitable. There are too many parameters to estimate for annual seasonality with weekly data. $\boxed{1}$

(g) *ARIMA(0,1,52).*

- Not suitable. This has no seasonality and too many parameters. $\boxed{1}$

(h) *ARIMA(2,0,2)(0,1,0)*$_{52}$.

- Suitable. The seasonal difference would account for the annual seasonility. $\boxed{1}$

(i) *Regression with Fourier terms for the annual seasonality.*

- Not suitable. The seasonality is not consistent enough between seasons. $\boxed{1}$

(j) *Dynamic regression with Fourier terms for the annual seasonality.*

- Suitable. This would probably work well as Fourier terms would capture the seasonality and the ARIMA model would capture the leftover dynamics with possibly a stochastic trend included. $\boxed{1}$

[Total: 20 marks]

— END OF SECTION B —

SECTION C

1. *Why can't you fit an ETS model to this data set? [3 marks]*

Too many parameters ($52 - 1$ for the weeks in a year) will need to be estimated, in order to estimate the seasonal component of the model. $\boxed{3}$

2. *You decide to use the STL decomposition shown in Figure 3, with ETS used for the seasonally adjusted data. Write down the equations for the fitted model, including the STL equation, and the equations for the components. [6 marks]*

(I am using the notation we have in the book for forecasting with decomposition)

STL decomposition

$$y_t = \hat{S}_t + \hat{A}_t$$

$\boxed{2}$

Seasonal naïve

$$\hat{S}_t = \hat{S}_{t-52} + \eta_t$$

$\boxed{2}$

ETS(M,N,N) on the seasonally adjusted component

$$\begin{aligned} \hat{A}_t &= \ell_{t-1}(1 + \varepsilon_t) \\ \ell_t &= \ell_{t-1}(1 + \alpha\varepsilon_t) \end{aligned}$$

$\boxed{2}$

3. *The residuals for the model are shown in Figure 4.*

What is causing the outlier at the start of 2022? [2 marks]

The unusual increase in the mortality rate at the beginning of 2022 which has been passed into the irregular component of the STL decomposition and hence in the residuals of the model. This is due to Covid (and in particular due to relaxing restrictions in Australia). $\boxed{2}$

Why are there no residuals for 2015? [2 marks]

There are no residuals as the seasonal naive method has been implemented for the seasonal component, and it needs at least a year of data to generate a forecast. $\boxed{2}$

Do you think the residuals would pass a white noise test? [2 marks]

The residuals look like white noise (no obvious patterns or autocorrelation left over) so yes I think they would pass the test. $\boxed{2}$

4. *The forecasts for the model are shown in Figure 5.*

What features of the data have not been captured by the model, and how does this affect the resulting forecasts? [3 marks]

The main strong feature not captured (and projected by the model) is the trend of the last couple of years. The Seasonally Adjusted component was projected with an ETS(M,N,N) hence no trend. Also the spike at the beginning of 2022 has not been captured and appears in the residuals. The model not capturing these results leads to wider prediction intervals. $\boxed{1}$ $\boxed{1}$ $\boxed{1}$

Discuss how the large prediction intervals would affect how this model would be used to inform policy decisions around mortality risks? [2 marks]

The large prediction intervals highlight/reflect high uncertainty. This should clearly be communicated to policy makers when presenting them with the forecasts. $\boxed{2}$

[Total: 20 marks]

— END OF SECTION C —

SECTION D

An alternative model for the same data is a dynamic regression model, fitted as follows.

1. *Write down the full model using backshift notation. [4 marks]*

$$y_t = \sum_{k=1}^{2} \left[\alpha_k \sin\left(\frac{2\pi k t}{52}\right) + \beta_k \sin\left(\frac{2\pi k t}{52}\right)\right] + \eta_t$$ $\boxed{2}$

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B)\eta_t = (1 + \theta_1 B)\varepsilon_t$$ $\boxed{2}$

2. *The number of pairs of Fourier terms was chosen by minimizing the AICc. Explain why this is a good way to choose a model being used for forecasting. [2 marks]*

Selecting models based on minimising the AIC (or AICc which corrects for small samples) assuming Gaussian residuals is asymptotically equivalent to minimising one-step time series cross validation MSE. Hence, this focuses on forecasting. $\boxed{2}$

3. *Comment on the model diagnostics shown in Figure 6 and the output below. [5 marks]*

- Time plot shows residuals are well behaved $\boxed{1}$
- ACF plot does not show any significant spikes (other than a couple at very long lags). $\boxed{1}$
- Histograms looks fairly normal $\boxed{1}$
- The couple of outliers at the beginning of 2022 are reflected in the times plot and also on the right tail of the histogram $\boxed{1}$
- LB test shows that we cannot reject the null of white noise $\boxed{1}$

4. *The point forecasts for 2023 are lower than for 2022. What feature of the model causes this? [2 marks]*

The model does not account for the trend in the mortality rate in the last couple of years and hence this is not reflected in the forecasts. $\boxed{1}$ Seasonality has been estimated using the Fourier terms throughout the sample. Hence, the projection reflects the average seasonality across the sample. $\boxed{1}$

5. *If you wanted to allow for the COVID-19 pandemic, how could you modify the model? [3 marks]*

You would want to include variables that capture the effect of the pandemic better than the current models. $\boxed{1}$ For example an indicator variable to capture the spike at the beginning of 2022 would help. $\boxed{1}$ Another covariate may be to include the number of Covid cases. $\boxed{1}$

6. *If you could obtain the actual weekly mortality rates for the first few months of 2023, how would you measure the accuracy of the point forecasts and the accuracy of the prediction intervals from this model? [4 marks]*

- For the point forecasts you would calculate the forecast errors and use an appropriate error measure (RMSE, MASE, RMSSE) to summarise these. $\boxed{2}$

- For the prediction intervals you could evaluate their coverage rate. For example you could evaluate whether your 80% prediction intervals contain 80% of the actual observations. (Alternatively, Winkler scores could be used instead of coverage rates.) $\boxed{2}$

[Total: 20 marks]

— END OF SECTION D —

SECTION E

You decide to compare the two models discussed in Q3 and Q4, along with a seasonal naïve benchmark. The following code uses a time series cross-validation with a 6 week test window.

1. *The minimal training set used in the time series cross-validation included 105 weeks. Why couldn't this be made any smaller? [2 marks]*

You need at least two years to estimate the seasonal component in the seasonal decomposition model. 2

2. *What do you conclude from the above output about the three models? [2 marks]*

The dynamic regression model with the Fourier terms included as predictors is the most accurate model with the seasonal naïve being the least accurate. 1 1

3. *Even before the COVID-19 pandemic, each year showed a different pattern from the previous year, especially in terms of the size of the peak. This is due to seasonal illness such as flu. What aspects of the STL+ETS and dynamic regression models allow for these differences between years? [4 marks]*

- The way both models are specified at the moment do not allow for these differences to be accounted for directly. 1
- The average seasonal variation is estimated across the years. The coefficients of the Fourier terms in the dynamic regression model are estimated across all the years and the seasonal component in the STL decomposition is fixed/periodic. 2
- Any variation in seasonal patterns will end up in the seasonally adjusted series (for STL+ETS) or in the ARIMA series (for dynamic regression). 1

4. *The accuracy statistics show that the mean error is smallest for the STL+ETS model, but the mean absolute error is smallest for the ARIMA model. Why would we prefer to select a model using MAE rather than ME? [3 marks]*

We base our evaluation of forecast accuracy on absolute or squared errors (in this case MAE). When calculating mean errors, negative and positive forecast errors will cancel out, hence we will get some indication of forecast bias. 1 If we do observe some forecast bias we can actually adjust our model (if required) to account for this. 1 In the case of both STL decomposition and regression models we know that our forecasts are based on residuals that are unbiased. 1

5. *The MASE and RMSSE accuracy measures use scaled errors. Why is scaling errors unnecessary in this analysis? [2 marks]*

In this analysis we are evaluating forecast errors for only one time series and hence there is no scaled differences that need to be take into account. 2

6. *The MAPE accuracy measure uses percentage errors. While percentage errors are meaningful in this example, sometimes they are not. When would the MAPE give meaningless or unhelpful results? [2 marks]*

MAPE is problematic when we have zero observations (in that case MAPE is undefined as we are dividing by zeros) or even observations that are close to zero (dividing by observations that are very close to zero can substantially skew our results leading to erroneous conclusions). Furthermore, percentage errors make no sense when the unit of measurement has no meaningful zero. 2

7. The output above shows the Winkler score for a 95% prediction interval. Explain in words how the Winkler score works in terms of the width of the prediction interval and the penalties that apply when the observation is outside the prediction interval. [5 marks]*

The Winkler score W_α first applies a penalty based on the width of the prediction interval where α is the level of significance considered. This is reflected by $(u_\alpha - \ell_\alpha)$. Hence, the wider the prediction interval the higher/worse the W_α. If the prediction interval contains the observed value then $W_\alpha = (u_\alpha - l_\alpha)$. If the observed value is outside the prediction interval an additional penalty is applied proportional to how far the observation is outside the interval. If the observation is below the lower bound the additional penalty is $\frac{2}{\alpha}(\ell_\alpha - y)$. If the observation is above the upper bound the additional penalty is $\frac{2}{\alpha}(y - u_\alpha)$. 1 1 1 1 1

[Total: 20 marks]

— END OF SECTION E —