Forecasting Exam 2022: 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. *Events like COVID-19 show that forecasting is never a good idea because the future is too unpredictable.*

- This is not true. [1]
- Often the future is relatively predictable if the future and past have similar conditions. [1]
- Events like COVID-19 can result in poor forecasts if the model has not allowed for such things. [1]
- But a model might allow for a major event like COVID-19, especially if a similar event has been observed before. [1]
- Prediction intervals are design to show uncertainty and in volatile environments, they will be very wide to indicate unpredictability. [1]

2. *Taking logarithms of the data is useful for stablizing the variance of a time series.*

- This is often true. [1]
- One way of dealing with heteroskedasticity is via transformations such as logarithms. [1]
- Other transformations that can be useful are the Box-Cox class of transformations. [1]
- These only work when the variance increases with the level of the series. [1]
- Logarithms have the additional advantage of being interpretable on a multiplicative scale. [1]

3. *Regression models are better than ARIMA models because the coefficients are more interpretable.*

- This is not true in a forecasting context. [1]
- Regression coefficients *are* more interpretable than ARIMA parameters. [1]
- But when forecasting, we are more interested in accuracy than in interpretability. [1]
- Without modification (such as dynamic regression), regression models do not allow for serial correlation. [1]
- Regression models also need the predictors to be forecasted. [1]

4. *The best forecasting model has white noise residuals.*

- This is usually true but incomplete. [1]
- White noise residuals is a sign that the forecasting model has captured the available information in a time series. [1]
- However, the model may be over-fitted. [1]

- Sometimes it is not possible to find a model with white noise residuals, but the forecasts can still be ok, especially if the autocorrelations are small (though possibly significant) [2]

5. *Choosing a model using the AICc is better than choosing a model on the basis of a test set because it involves all the data.*

- This is usually true when comparing a model within the same family of models. [1]
- However, you can't compare AICc values across model classes (such as ETS and ARIMA), and you can't compare AICc values across ARIMA models with different levels of differencing. [2]
- The AICc also assumes Gaussian residuals. It can be misleading if this is not true. [1]
- When the AICc can't be used, a better method than a test set is time series cross-validation. [1]

6. *The MAPE is the best accuracy measure because it is easy to understand and is independent of the scale of the data.*

- This is not true. [1]
- Yes, the MAPE is easy to understand and is independent of the scale of the data. [1]
- But it is not optimal for the mean (or any other well-defined aspect of a distribution). (GA: I did not cover this - so will be more lenient here) [1]
- It is also not helpful when there are zeros or small values in the data, or when the data do not have a natural zero. [1]
- For measuring point forecast accuracy, the RMSE or RMSSE is usually better. (GA: I did not cover this - so will be more lenient here) [1]

[Total: 20 marks]

— END OF SECTION A —

SECTION B

1.
- the time plot shows an increasing trend [1]
- the variance grows with the level of the series [1]
- there are two types of seasonality: annual and weekly [1]
- the annual seasonality shows a low period at the end/beginning of each year, and peaks around March/April and October/November. [1]
- the weekly seasonality shows the lowest days are weekends, with the maximums on Tuesday–Friday. [1]
- there are a couple of large positive outliers (in 2018 and 2021). [1]

2.
- a logarithm is used to make the variance more stable [1]
- the yearly `window` of 9 is actually large enough because there are only 5 years of data, so all years are included in each window. A larger window wouldn't make much difference; a very small window might allow the annual seasonality to change from year to year. [1]
- the weekly `window` of 99 is small enough to allow change over time because only 99 weeks (roughly two years) are included in each window. For periodic weekly seasonality, it would have to be much larger; a smaller window would allow the weekly seasonality to change more rapidly. [1]
- the trend `window` of 99 is small enough to allow substantial wiggliness in the trend because it is only including 99 days in each window. A much larger window would force the trend to be linear; a much smaller window would force the trend to be more wiggly. [1]

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.

```
- not ok. It would not allow for the weekly seasonality.\marks{1}
```

(b) Seasonal naïve method using weekly seasonality.

```
- not ok. It would not allow for the annual seasonality.\marks{1}
```

(c) An STL decomposition on the log transformed data combined with an ETS to forecast the seasonally adjusted component, and seasonal naïve methods for both seasonal components.

```
- this would probably work well.\marks{1}
```

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

```
- not ok. It would not allow for the annual seasonality.\marks{1}
```

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

```
- not ok. It would not allow for the annual seasonality.\marks{1}
```

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

- not ok. There are too many parameters to estimate and it would not allow for the weekly seasonalit

(g) ARIMA(2,2,2)$(0,0,0)_7$ applied to the log transformed data.

- not ok. This has no seasonality.\marks{1}

(h) ARIMA(0,1,1)$(0,1,1)_7$ applied to the log transformed data.

- not ok. It would not allow for the annual seasonality.\marks{1}

(i) Regression with time and Fourier terms for both weekly and annual seasonality.

- not ok. The trend is not linear.\marks{1}

(j) Dynamic regression with Fourier terms for the annual seasonality and a seasonal ARIMA model to handle the weekly seasonality and other dynamics.

- this would probably work well.\marks{1}

[Total: 10 marks]

— END OF SECTION B —

SECTION C

1. *Write down the observation and state equations for the model, specifying which is the observation equation, what parameters have been optimized, and explaining why this particular model was chosen.*

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

2

- The first equation (LHS y_t) is the observation equation. The others are state equations. 1
- Parameters optimized: $\alpha, \beta, \gamma, \ell_0, b_0, s_0, s_{-1}, \dots, s_{-5}$ [1 mark if included 7 seasonal states. OK to omit a different seasonal state.] 2
- Model chosen because it minimized AICc (GA: The "why" here may lead them to comment on the features of the model, e.g., it will account for the changing seasonality, local trend, etc. Should we change the "why" to "how" has the model been chosen?) 1

2. *Figure 5 shows the components of the model. Explain what has been plotted, and how these plots relate to the equations shown earlier. Explain how the values of `beta` and `gamma` shown earlier correspond to features of these plots.*

- The plots show (in order) y_t, ℓ_t, b_t, s_t and ε_t plotted against t. 2
- i.e., the LHS of each equation shown in first four plots, and ε_t in last plot. 1
- Small β indicates slope changes very slowly (over four years it only changes from 0.145 to 0.115). 1
- Larger γ indicates weekly seasonality changes more rapidly over time (fourth panel). 1

3. *How have the annual and weekly seasonalities been handled by this ETS model?*

- Annual seasonality ignored in the model. 1
- Annual seasonality gets picked up in the level component. 1
- Weekly seasonality handled via s_t seasonal component. 1

4. *Comment on the large residuals seen at the end of each year, and the dip in the level at the end of each year. What is causing these?*

- Both are caused by the annual seasonality. 2

5. *Do you expect this model to produce good forecasts for the next 3 weeks? What about for the next 12 months? Explain.*

- It should be ok for the next few weeks because the end of the series is relatively flat so the level projection will be ok, and the seasonality won't change much over a few weeks. 2
- 12 month forecasts will be very poor because it won't forecast the annual seasonality at all. 2

[Total: 20 marks]

— END OF SECTION C —

SECTION D

It is decided to fit a dynamic regression model to the data, with Fourier terms to handle the annual seasonality, and a seasonal ARIMA error to handle the weekly seasonality.

1. *Write down the model using backshift notation.*

$$\log(y_t) = \sum_{k=1}^{4}\left[\alpha_k \sin\left(\frac{2\pi kt}{365}\right) + \beta_k \cos\left(\frac{2\pi kt}{365}\right)\right] + \eta_t$$ 2

$$(1-\phi_1 B)(1-\Phi_1 B^7 - \Phi_2 B^{14})(1-B^7)\eta_t = (1+\theta_1 B + \theta_2 B^2)(1+\Theta_1 B^7)\varepsilon_t$$ 2

where $\varepsilon_t \sim \text{NID}(0, \sigma^2)$ and 1

$\phi_1 = 0.886$, $\Phi_1 = 0.160$, $\Phi_2 = 0.025$, $\theta_1 = -0.121$, $\theta_2 = -0.175$, $\Theta_1 = -0.905$, $\sigma^2 = 0.009$. 1

2. *Comment on the model diagnostics shown in Figure 6 and the output below. How might the model be improved? Do you think the resulting forecasts and prediction intervals will be reliable?*

- There are outliers at the end of each year suggesting the annual seasonality has not been dealt with adequately. 1
- The ACF shows several significant autocorrelations, but these are very small and probably won't have an effect. 1
- The Ljung-Box test is significant, showing the residuals are not white noise. But this is probably not a problem as the autocorrelations are small. 1
- The model can be improved by handling the annual seasonality better – either increasing the number of Fourier terms, or including some end-of-year dummy variables. 1
- Resulting forecasts not great at end of year, but otherwise should be ok. 1

3. *Figure 6 shows forecasts with prediction intervals for the next four weeks, along with the data from 2022. The forecasts appear to have no trend. Why is that? If you wanted to include a* local* trend, how would you modify the model?*

- No trend because there is no trend in the model, and there is only one level of differencing ($D = 1$). 2
- Could add local trend with additional differencing, setting $d = 1$. 1

4. *The one-step-ahead forecast is 35.0 pageviews. Give a 95% prediction interval assuming Gaussian innovation residuals.*

- On log scale: $\log(35.0) \pm 1.96\sqrt{0.009} = [3.37, 3.74]$. 2
- So on original scale, take exponentials: [29.07, 42.18] 2

5. *It is thought that pageviews will be higher during teaching semesters. How would you modify this model to allow for a semester effect?*

- Add dummy variable as an additional regressor for days within semester. 2

[Total: 20 marks]

— END OF SECTION D —

SECTION E

1. *You decide to compare three different models on this data set: (a) the ETS model from Section C; (b) the dynamic regression model from Section D; and (c) the STL decomposition shown in Figure 4 with ETS applied to the seasonally adjusted data, and seasonal naive methods applied to both seasonal components.*

   The following code uses a test set of the last 4 weeks to compare the three models, along with two benchmark methods.

   What do you conclude from the above output about the five models? Explain the two accuracy measures used. Why is the naive method so bad?

- RMSE is root mean squared error, defined as $\sqrt{\frac{1}{H}\sum_{h=1}^{H}(y_{T+h} - \hat{y}_{T+h|T})^2}$. 1
- MAPE is mean absolute percentage error, defined as $\frac{1}{H}\sum_{h=1}^{H}|y_{T+h} - \hat{y}_{T+h|T}|/\hat{y}_{T+h}$. 1
- Dynamic regression is best using either accuracy measure. 1
- Naive is particularly bad because it doesn't allow for any form of seasonality. 1

2. *An alternative approach to comparing the forecast accuracy of models would be to use time series cross-validation. Explain the concept of time series cross-validation. You may use an annotated diagram.*

- TSCV involves fitting models to multiple training sets 1
- If available data is $y_1, \dots, y_T$, then training sets are expanding of the form $y_1, \dots, y_t$, where $t = p, p+1, \dots, T$. 1
- Initial training set of size p where p chosen to ensure a model can be estimated. 1
- Evaluation on y_{t+h} for a specific h and all t. 1
- Good diagram 2

3. ***ETC3550 students only***: *You decide the* `snaive` *model is good enough even though it is not as accurate as the first three models. Let the observations be $y_1, \dots, y_T$ and the residual variance be denoted by σ^2. Write down the forecast distribution for an h-step forecast. What assumptions have you made?*

- Mean $\hat{y}_{T+h|T} = y_{T+h-7(k+1)}$ where $k =$ integer part of $(h-1)/7$. 1
- Var $\hat{\sigma}_h^2 = (k+1)\hat{\sigma}^2$ where $\hat{\sigma}^2$ is residual variance. 1
- Distribution: $\text{N}(\hat{y}_{T+h|T}, \hat{\sigma}_h^2)$ 1
- Assumed Gaussian residuals, homoscedastic error, uncorrelated errors. 2

4. ***ETC5550 students only***: *Someone proposes a fancy new forecasting method that uses some predictor variables such as the number of students in each university that is using an OTexts book as a recommended text, and the relative wealth of the countries they live in. This new method reduces the cross-validated one-step RMSE by 2%, but takes about a day to estimate the forecasts. Would you recommend using the new method? Explain your reasons.*

- Probably not. 1
- The additional time and energy to compute forecasts does not seem worth it. 1

- Also, the predictors will need to be forecast, and that information may not be available in advance, especially student numbers. 2
- Even collecting the historical data on student numbers in each university is a big task. 1

5. *OTexts also needs forecasts of the maximum monthly traffic that could arise, in order to make sure their internet server will cope with the extreme demand. They want to choose an internet plan that allows for a maximum of P pageviews per month, and they are happy to allow a 1% chance of being above this level. Explain how you would choose P based on forecasts over the next 12 months.*

- Aggregate data to monthly. 1
- Fit a model to the monthly data — something like an ETS or an ARIMA should be ok. 1

Method 1

- Compute 12 month forecasts and find the 12 future forecast distributions. 1
- Set P to be the maximum of the twelve 99% quantile values. 1
- This does not actually give a 1% chance of being above level, either over a year or over a month. A mark awarded if the student recognizes this, or tries to address it in some way. 1

Method 2

- Simulate many future sample paths of length 12 from fitted model, and choose P to be the 99% quantile over all sample paths. 2
- That sets the probability of 1% per year. Need a different quantile if you want the probability of being above threshold over a different time period. 1

[Total: 20 marks]

— END OF SECTION E —