Forecasting Exam 2021: 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. No marks will be given without any justification.)*

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. *Narrower prediction intervals are more informative and should always be preferred.*

- This is not true. [1]
- We want our prediction intervals to have as close as possible to the correct coverage, without that defining whether they are narrower or wider. [2]
- Usually prediction intervals will have lower coverage than what they aim for, i.e., they will be narrower, as they do not account for all sources of uncertainty, e.g., they do not account for the uncertainty of the model choice and they assume that the model is the correct model. [1]
- A wide prediction interval is informative in terms of signaling the large uncertainty associated with the prediction. [1]

2. *The AICc should always be used to select models for forecasting.*

- This is almost true. [1]
- The AICc is very useful and we prefer to use it when possible, as it uses all the sample and is asymptotically equivalent to minimizing one-step time series cross validation MSE assuming Gaussian residuals. [2]
- However it is limited and cannot be used when comparing across different classes of models or even sometimes within the same class (for example ARIMA models with different orders of differencing, or models with different transformations). [2]

3. *An ETS model for Holt's linear trend method, is a generalisation of an ETS model for simple exponential smoothing. It should therefore always be preferred as it will produce better forecasts.*

- This is not true. [1]
- Yes ETS(A,A,N) is a generalisation of ETS(A,N,N). [1]
- However the generalisation involves estimating an extra smoothing parameter and an extra initial state for the slope. [1]
- An information criterion, such as the AICc, will penalise the fit of the model for the extra degrees of freedom (extra estimation uncertainty) and guide us in choosing between the two models. [2]

4. *The trouble with forecasting is that it assumes the patterns in the past will continue in the future.*

- This is true. 1
- Any forecasting model is based on historical data and attempts to extrapolate model into the future. 2
- However models do not assume things stay the same. ETS models, for example, allow all components to change over time. ARIMA models allow for non-stationarity by modelling changes rather than levels. 1
- The issue is whether the way things change after the training data has been captured in the model. 1

5. *An ARIMA model with uncorrelated residuals will always produce accurate forecasts.*

- This is not true. 1
- A good forecasting model should have uncorrelated residuals, but uncorrelated residuals does not mean the model is good for forecasting. 2
- For example, a model may be overspecified by including unnecessary higher order terms. These will make the forecasts worse than they need to be, although the residuals will remain uncorrelated. 2

6. *Regression models with Fourier terms should always be used to model seasonality.*

- This is not true. 1
- Fourier terms are particularly useful when dealing with long seasonal periods. For example, for weekly data where something like 52 dummies would be required, using fewer Fourier terms could be advantageous. 2
- Using all $m-1$ Fourier terms for quarterly or monthly data would be equivalent to using dummy variables. 1
- Also regression models assume that seasonality is deterministic, i.e., they do not allow for it to change over time. Alternative specifications, such as ETS models, decomposition methods, ARIMA models may be preferred if such an assumption is not suitable. 1

[Total: 20 marks]

— END OF SECTION A —

SECTION B

1. Marks are allocated for the six main features, overall trend (1), cycles (peaks-troughs) (1), seasonality (2), outliers (1).

- the time plot shows a long-term/global upward trend, hence the number of births have overall grown over time. 1
- the time plot shows that the series has cyclical features, peak around 1990-1991, trough around 2001-2002 and another downturn starting around 2016. 1
- the time plot shows constant seasonal variation throughout the sample, i.e., additive seasonality. 1
- although the seasonal plot is a little messy there seems to be a drop in February and an increase in March followed by some minor fluctuations in the months that follow. This is possibly because of the fewer days especially in February (28 and 29) and some of the other months. Clearly a substantial drop in November every year with October being the peak. These are also reflected in the subseries plot with February and November showing the lowest average number of births. 1
- the plot shows a few outliers at the end of 2000 and beginning of 2001. These are clearly shown in the seasonal plot. 1

2.

- Figure 4 shows an STL decomposition, where the lower three panels show trend-cycle, seasonal and remainder components.

- It seems that the default setting of `season(window=13)` allows the seasonal component to change quite a lot. I might experiment with increasing this window to possibly stabilise this a little. 1

- The outliers in Dec 2000 and Jan 2001 are shown in the remainder series and also a large outlier seems to show up in the remainder series towards the end of the sample. I would experiment with setting `robust=TRUE` so that such outliers are better absorbed into the remainder series especially if I am only interested in forecasting the trend component. I suspect this will make a difference to the end of the trend. 1

- Figure 5 shows forecasts for the next two years using the STL decomposition. The seasonal component is projected using the seasonal naïve method and the seasonally adjusted series is projected using a random walk hence naïve forecasts. The two are then combined. 1

- The prediction intervals seem to be very wide. These are possibly appropriate given the cyclical feature of the data. 1

- I would possible experiment with better forecasting the seasonally adjusted component give then downward trend towards the end of the series. 1

3. (a) *Seasonal naïve method.*

   Suitable. We are not sure when the series will break out of the downturn, hence the downward trend may be a local trend. 1

(b) *An STL decomposition combined with the drift method to forecast the seasonal adjusted component.*

Not suitable. The drift will project the global trend although the series is on a downturn local trend. [1]

(c) *An STL decomposition on the log transformed data combined with an ETS to forecast the seasonally adjusted component.*

Not suitable. The seasonal component is additive no need for a transformation. [1]

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

Suitable, both additive seasonality and a damped trend would be appropriate as we expect the down turn to flatten out. [1]

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

Possibly suitable. Additive component required. No trend may be suitable given the downturn. [1]

(f) *ETS(A,Ad,M).*

Not Suitable as seasonality is additive but also an unstable model. [1]

(g) *ARIMA(1,1,4).*

Not suitable. Data is seasonal. [1]

(h) *ARIMA(3,0,2)(1,1,1)*$_4$.

Not suitable the data is monthly. [1]

(i) *ARIMA(1,0,2)(2,1,0)*$_{12}$*}.*

Suitable if the seasonal difference is enough to reach stationarity. Of course you will need to check the residuals. [1]

(j) *Regression with time and Fourier terms.*

Not suitable. Fourier terms will be ok for seasonality however the trend is clearly non-linear and we need something to account for that. Something like a piecewise trend will probably be better with knots around 1990, 2000 and 2016. [1]

[Total: 20 marks]

— END OF SECTION B —

SECTION C

1.

- `fit_ETS` is a mable (model table) object containing three ETS models for the birth series. [1]
- The models differ in their trend component: `ets_N` has no trend, `ets_A` has an additive trend and `ets_Ad` has an additive damped trend component. [1]
- The tibble shows the coefficients of each model and the initial states. Columns 2 shows the extra smoothing parameter β and the initial value for the slope and column 3 shows the extra damping parameter ϕ. [1]

2. The plot shows the data in the first panel and below that the estimated components of the estimated ETS models.

- The level for each model is adjusting over time, with fairly similar smoothing parameters α of 0.468, 0.327 and 0.299, following closely the cyclical nature of the data. [1]
- The slope panel shows only two series with the ets_N model not including a trend component. The β for the other two models is fairly low (close to 0) however it reflects some change in the slope (although this is fairly small — note the scale difference between the level and the slope components) adjusting to cyclicity of the series and we also see a slight difference in the slope between the two models with the ets_Ad also including a damping parameter. [1]
- In contrast to the STL decomposition the seasonal components of the three models do not change much with very low γ smoothing coefficients. [1]
- The remainder component reflects the outliers of Dec 2000 and Jan 2001 [1]

3.

- Looking at the MSE the extra parameters for the ets_A and ets_Ad models help with improving their fit. [1]
- Based on AICc values, either of the two trended models would be better than the model with no trend. But there is almost no difference between the two. I'd choose the simpler (ets_A) model, but either is ok. [1]

4. The estimated `ets_N` model is:

$$\begin{aligned} y_t &= \ell_{t-1} + s_{t-m} + \varepsilon_t \\ \ell_t &= \ell_{t-1} + 0.468\varepsilon_t \\ s_t &= s_{t-m} + 0.0002\varepsilon_t \end{aligned}$$

$\varepsilon_t \sim N(0, 0.0180)$. [3]

5. Point forecasts:

$$\begin{aligned} \hat{y}_{T+1/T} &= \hat{y}_{Oct,2019} = 5.86 + 0.276 = 6.14 \\ \hat{y}_{T+4/T} &= \hat{y}_{Jan,2020} = 5.86 - 0.0146 = 5.85 \\ \hat{y}_{T+12/T} &= \hat{y}_{Sep,2020} = 5.86 + 0.142 = 6.00 \\ \hat{y}_{T+13/T} &= \hat{y}_{Oct,2020} = 5.86 + 0.276 = 6.14 \end{aligned}$$

[3]

6. Prediction interval:

$$\text{Oct 2019} : 6.14 \pm 1.28 * \sqrt{0.0180} = (5.97, 6.31)$$ 3

7. • Although both sets of forecasts are flat there is a stark difference in the prediction intervals with the RW model generating much wider prediction intervals. 1

Longer term forecasts:

- The forecasts from ets_N will be flat with the seasonal component repeated.
- The forecasts from the ETS_A and ETS_Ad models will be trended. From Figure 6, $b_T > 0$ for ETS_A and $b_T < 0$ for ETS_Ad. So ETS_A forecasts will trend upward, while ETS_Ad forecasts will trend downard and then flatten out. 1

[Total: 20 marks]

— END OF SECTION C —

SECTION D

1.
   - The original birth series has been seasonally differenced and that is what we observe in Figure 7. 1
   - The seasonally differenced series seems to be stationary - I would be happy to model based on this ACF and PACF. 1
   - The KPSS test confirms my intuition as the NULL of stationary series cannot be rejected 2
2.

- Seasonal component. ACF shows 1 spike, PACF exponential decay. Hence, I choose PDQ(0,1,1) 2
- Non-seasonal. PACF shows 5 spikes, ACF decaying. Hence I choose pdq(5,0,0) 2

3. We expect something like the following output.

```
## Series: count
## Model: ARIMA(5,0,0)(0,1,1)[12]
##
## Coefficients:
##          ar1     ar2     ar3      ar4     ar5     sma1
##       0.3708  0.2079  0.2147  -0.0319  0.2140  -0.8844
## s.e.  0.0458  0.0488  0.0490   0.0486  0.0462   0.0329
##
## sigma^2 estimated as 0.01773:  log likelihood=271.73
## AIC=-529.46   AICc=-529.22   BIC=-500.47
```

   - Correct model and pasted estimation output. 1

```
## # A tibble: 1 x 3
##   .model                                     lb_stat lb_pvalue
##   <chr>                                        <dbl>     <dbl>
## 1 ARIMA(count ~ pdq(5, 0, 0) + PDQ(0, 1, 1))    30.1  0.000417
```

   - Comment on the residual plot 1
   - Ljung-Box for lag 14 with dof=5 rejects the null of WN at 5% level of significance. 1

4.
   - Model is selected by the stepwise search algorithm of Hyndman and Khandakar 1
   - The residuals look reasonable — there are some significant spikes though. I may try and switch the stepwise process off to see whether I can get a better model. 1
   - I would choose the model from Q3 (better AICc and better resids) 1

5.
   - the model is projecting the trend in the end. This has been picked up with the AR(1) parameter being close to 1 (so almost a second difference which would lead to a local trend). 1
   - only one difference will converge to a constant 1

6. Write out the model

$$(1-\phi_1 B)(1-\Phi_1 B^{12}-\Phi_2 B^{24})(1-B^{12})y_t = (1+\theta_1 B)e_t$$

1

$$[1-\phi_1 B-(1+\Phi_1)B^{12}+\phi_1(1+\Phi_1)B^{13}-(\Phi_2-\Phi_1)B^{24}+\phi_1(\Phi_2-\Phi_1)B^{25}+\Phi_2 B^{36}-\phi_1\Phi_2 B^{37}]y_t = (1+\theta_1 B)e_t$$

2

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + (1+\Phi_1)y_{t-12} - \phi_1(1+\Phi_1)y_{t-13} + (\Phi_2-\Phi_1)y_{t-24} - \phi_1(\Phi_2-\Phi_1)y_{t-25} \\ &\quad - \Phi_2 y_{t-36} + \phi_1\Phi_2 y_{t-37} + e_t + \theta_1 e_{t-1} \end{aligned}$$

1

[Total: 20 marks]

— END OF SECTION D —

SECTION E

1.
   - Each model is a dynamic harmonic regression with a piece wise linear trend with knots at Jan 1991, Jan 2001 and Jan 2016. [1]
   - I would select `pw6` as it has the lowest AICc value. [1]
2.
   - This model is equivalent to fitting a dynamic regression with seasonal dummies. [1]
3. $y_t = \beta_0 + \beta_1 t + \beta_2(t-\tau_1)_+ + \beta_3(t-\tau_2)_+ + \beta_4(t-\tau_3)_+ + \sum_{k=1}^{6}\left[\alpha_k \sin\left(\frac{2\pi kt}{m}\right) + \gamma_k \cos\left(\frac{2\pi kt}{m}\right)\right] + \eta_t,$ [2]

$$\text{where} \qquad (1-\phi_1 B - \phi_2 B^2)(1-\Phi_1 B^{12})\eta_t = (1+\theta_1 B)(1+\Theta_1 B^{12} + \Theta_2 B^{24})\varepsilon_t,$$

[2]

$\varepsilon_t \sim N(0, \sigma^2)$, [1]

τ_1 corresponds to Jan 1991, τ_2 corresponds to Jan 2001, and τ_3 corresponds to Jan 2016. [1]

- β_1 is the trend prior to Jan 1991. [1]
- $\beta_1 + \beta_2$ is the trend between Jan 1991 and Jan 2001. [1]
- $\beta_1 + \beta_2 + \beta_3$ is the trend between Jan 2001 and Jan 2016. [1]
- $\beta_1 + \beta_2 + \beta_3 + \beta_4$ is the trend after Jan 2016. [1]

4.
   - Figure 10 shows the fitted values together with the forecasts generated by model `pw6`. The fitted values show that the knots have been well placed capturing the cyclical feature (peaks and troughs) in the series. [1]
   - The forecasts project the final estimate of the trend. [1]
   - The residuals seem to be well behaved in terms of their mean and distribution, however there are the few outliers that we noticed in the previous questions. We could fit dummies to capture and remove those. [1]
   - The ACF shows a few significant spikes which may affect the prediction intervals generated however the values are very low and the spikes are not at small lags. I could try to increase the orders of the ARMA model for the errors but I would be satisfied even with these residuals. [1]
5. Any reasonable discussion along the lines below is fine. [3]

- The model projects a downward trend — this is very sensitive to where the last knot is placed - in this sense I need to be cautious. In any case I think these forecasts would be ok for the short run but nothing beyond that.
- I would prefer forecasts from the STL/RW, ETS_N, ETS_Ad for medium to longer term forecasts as they will flat in the longer term. So will the ARIMA ones.
- The uncertainty shown by the STL/RW forecasts is possibly too large, but the uncertainty shown by some of the other models seems too small. It is challenging to predict the next upturn of the series.

[Total: 20 marks]

— END OF SECTION E —

SECTION F