ETC3550 Exam 2018: solutions

QUESTION 1

*Write up to one quarter of a page each on any **four** of the following topics. (You may 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.

(a) *Prediction intervals are unnecessary because managers just want point forecasts.*

This is not true. While many people say they just want point forecasts, they are misleading without any statements of uncertainty. Prediction intervals are one simple way of quantifying the uncertainty in forecasts, and prevent over-interpretation of minor variation in the point forecasts. Prediction intervals also allow for better planning when the forecasts are used to make decisions for the future. 5

(b) *Whether we use a naïve approach, a decomposition, ETS or ARIMA models for forecasting, we always need to transform our data.*

This statement is not true. Some decompositions can be multiplicative, e.g. classical, X11, and also ETS models can cope with multiplicative. We need to take a transformation to stabilise the variance if we are using ARIMA models. This will also help if we were using a naïve approach. 5

(c) *Simple exponential smoothing should only be used for a series with a constant underlying level.*

This is not true. SES should not be used where there is a trend or seasonality. However, the underlying level does not need to be constant. The series is allowed to wander around as SES allows for a unit root. We also showed that SES is equivalent to an ARIMA(0,1,1). So the time series does not need to be stationary. 5

(d) *All three information criteria: AIC, AICc and BIC are useful and they can potentially choose a different model. We prefer to use the AICc forecasting.*

This is true. AIC applies the smallest penalty and is asymptotically equivalent to minimising the tsCV for 1-step-ahead - hence something based on a forecasting criterion. AICc applies a bias correction to the AIC which may make a difference in small samples - for large samples these will be identical. The BIC applies the largest penalty and therefore will always choose the most parsimonious model as long a $\log(T) > 2$, i.e., $T > 8$. The BIC can be useful in other contexts as it is consistent for large T however that relies on a true model existing which is never the case. 5

(e) *The 95% prediction interval for a h-step-ahead naïve forecast is given by* $\hat{y}_{T+h} \pm 1.96\sqrt{h\sigma^2}$.

This is true. Assuming $e_t \sim N(0, \sigma^2)$, forecasts from the naïve method are given by $\hat{y}_{t/t-1} = y_{t-1}$; $e_t = y_t - \hat{y}_{t/t-1} = y_t - y_{t-1} \Rightarrow y_t = y_{t-1} + e_t$

$$
\begin{aligned}
y_{T+1} &= y_T + e_{T+1} \\
y_{T+2} &= y_{T+1} + e_{T+2} = y_T + e_{T+1} + e_{T+2} \\
&\vdots \\
y_{T+h} &= y_T + e_{T+1} + e_{T+2} + \cdots + e_{T+h} \\
\hat{y}_{T+h} = \mathrm{E}(y_{T+h|T}) &= y_T + \mathrm{E}(e_{T+1|T}) + \mathrm{E}(e_{T+2|T}) + \cdots + \mathrm{E}(e_{T+h|T}) \\
&= y_T \\
\mathrm{V}(y_{T+h|T}) &= 0 + \mathrm{V}(e_{T+1|T}) + \mathrm{V}(e_{T+2|T}) + \cdots + \mathrm{V}(e_{T+h|T}) \\
&= h\sigma^2
\end{aligned}
$$

Hence prediction interval $\hat{y}_{T+h} \pm 1.96\sqrt{h\sigma^2}$ $\boxed{5}$

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

This is not true. 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. Furthermore, scenario based forecasting is very useful for decision making. $\boxed{5}$

[Total: 20 marks]

— END OF QUESTION 1 —

QUESTION 2

(a) Figure 1 is a time plot. The shows a strong downward trend since 1990 (a fantastic result in terms of smoking). It seems that seasonality decreases towards the end of the series and it really diminishes in the last few years. If it is not a recording error or some change in definition by the ABS it seems that the taking up smoking in the 4th quarter (something we see in the next plot) and then quitting in the new year (Q1) is becoming less of a trend. 2

Figure 2 is a seasonal subseries plot which helps us identify any changes in seasonality over time especially for each particular season. This shows that the acerage of Q4 is the highest and hence the spikes in the time series correspond to the Q4. 2

(b) Yes, the series needs a transformation as the variation is decreasing as the level of the series is dropping. 1

We see a log and a Box Cox transformation with a lambda close to 0 — so they are both doing a similar thing. I would just go with the log in this case as it is easier to explain/communicate. 1

(c) The function first applies an log transformation, followed by an STL decomposition. 1

Then using the ETS(A,A,N) model selected by AICc, it generates forecasts for the trend. These are added to the seasonal naïve forecasts for the seasonal component, to give forecasts of the logged data. Finally, these are back-transformed to give forecast son the original scale. 3

(d) A. *Seasonal naive model.* Arguably suitable, the trend may be deemed to flatten towards the end. We'll take both yes and no answers with clear justification. 1

B. *Drift method plus seasonal dummies.* Potentially suitable - better with a log transformation - do not penalise if they do not say this. 1

C. *Holt's trend method.* Not suitable, data are seasonal. 1

D. *Holt-Winters multiplicative damped trend method.* Suitable. 1

E. *ETS(A,A,M).* One of the combinations that is not allowed. 1

F. *ETS(M,*A_d*,M).* Suitable for both accounting for changing seas and dampening of the trend. 1

G. *ARIMA(0,1,3).* Not suitable - no seas component. 1

H. *ARIMA(0,1,2)(1,1,0)*$_4$. Possibly suitable. 1

I. *ARIMA(0,0,1)(2,0,0)*$_4$. Not suitable you would need at least one lot of differencing. 1

J. *Regression with time and Fourier terms.* Trend not linear and seasonality changing. 1

[Total: 20 marks]

— END OF QUESTION 2 —

QUESTION 3

(a) The second specification `fit2.ets` includes a log transformation and therefore the model selected has all additive components. $\boxed{1}$

The estimated coefficients in both cases seem to very similar. In both models all the change is happening through the level with beta close to zero for fit2. The gamma coefficient seems quite large so lots of change is happening in the seasonal component. $\boxed{2}$

The second model has a trend component which will make a difference in the forecasts generated. $\boxed{1}$

(b) These are plots of the components. We can see the rapidly changing level and seasonal components. $\boxed{2}$

It seems for `fit1.ets`, some of the trending behaviour in the time series may have leaked into the seasonal component which is not only rapidly changing but also seems to be trending. $\boxed{2}$

(c)

$$
\begin{aligned}
y_t &= \ell_{t-1} s_{t-4}(1+\varepsilon_t) \\
\ell_t &= \ell_{t-1}(1+0.709\varepsilon_t) \\
s_t &= s_{t-4}(1+0.291\varepsilon_t), \qquad \varepsilon_t \sim N(0, 0.031^2)
\end{aligned}
$$

$\boxed{3}$

(d) Start with the residual plot. There seems some autocorrelations left in the residuals with some sequences of positive residuals and negative residuals $\boxed{1}$

The ACF shows no significant spikes other than a spike at lag 7 but some close to the critical value boundaries - but all in all if this was the only thing I would look at I could not distinguish these resids from white noise. However as we have a multiple testing problem hence the Ljung-Box test may be better. $\boxed{1}$

$H_0: \rho_1 = \cdots = \rho_9 = 0$, `fitdf=6`=2(smoothing coef)+1(level) +3(seasonals). This rejects the Null of WN at a 5% so not as convincing as the ACF. $\boxed{1}$

My point forecasts will be ok but possibly PI coverage will be incorrect. $\boxed{1}$

(e)

$$
\begin{aligned}
\hat{y}_{T+2} &= 1.2404 * 0.91493 = 1.1348 \\
\hat{y}_{T+4} &= 1.2404 * 0.95164 = 1.1804 \\
\hat{y}_{T+8} &= \hat{y}_{T+4}
\end{aligned}
$$

$\boxed{3}$

(f) Any sensible comment about back-transformed point forecasts are medians not means and need to do an adjustment if we want the means. $\boxed{2}$

[Total: 20 marks]

— END OF QUESTION 3 —

QUESTION 4

(a)
- First set: ACF and PACF of the raw data. The large slow decaying spikes in the ACF show that the data is not stationary (this is due to the prominent trend) and also seasonal spikes are seen. $\boxed{1}$
- Second set: ACF and PACF of the first differences. The large spikes show that the first difference is non-stationary with strong seasonality.Seasonal differencing is required. $\boxed{1}$
- Third set: ACF and PACF of the seasonal differences. ACF large spikes and rapid decaying suggests stationarity. Not much seasonality left. $\boxed{1}$
- Fourth set: ACF and PACF of seasonal and first order differences. Series clearly stationary with seasonality seen in lag 4 spikes in both plots. $\boxed{1}$

(b) Using the third set:

- Seasonal significant spikes at PACF 1 suggests ARIMA(p,0,q)(1,1,0). $\boxed{2}$
- Non-seasonal significant spikes at PACF 1 suggests ARIMA(1,0,0)(1,1,0). $\boxed{2}$
- Accept any reasonable justification for other models.

Or using the fourth set.

- Seasonal sign spikes at ACF:4, PACF: 4,8,12 - (p,1,q)(0,1,1) $\boxed{2}$
- Non-seasonal largish spikes at ACF:1. PACF: 1 - either (1,1,0)(0,1,1) or (0,1,1)(0,1,1) or (0,1,0)(0,1,1) $\boxed{2}$
- Accept any reasonable justification for other models.

(c)

$$(1-\phi B)(1-\Phi B^4)(1-B^4)(y_t-\mu t)=\varepsilon_t$$
$$(1-\phi B-\Phi B^4+\phi\Phi B^5)(y_t-y_{t-4}-4\mu)=\varepsilon_t$$
$$y_t-\phi y_{t-1}-\Phi y_{t-4}+\phi\Phi y_{t-5}-y_{t-4}+\phi y_{t-5}+\Phi y_{t-8}-\phi\Phi y_{t-9}=4\mu(1-\phi-\Phi+\phi\Phi)+\varepsilon_t$$
$$y_t=4\mu(1-\phi-\Phi+\phi\Phi)+\phi y_{t-1}+(1+\Phi)y_{t-4}-\phi(1+\Phi)y_{t-5}-\Phi y_{t-8}+\phi\Phi y_{t-9}+\varepsilon_t$$
$$y_t=c+\phi y_{t-1}+(1+\Phi)y_{t-4}-\phi(1+\Phi)y_{t-5}-\Phi y_{t-8}+\phi\Phi y_{t-9}+\varepsilon_t$$

where $y_t=\log(\text{CTC})$ and ε_t is white noise with mean 0 and variance 0.000912^2.

The parameters are $\mu=-0.006$, $\phi=0.766$, $\Phi=-0.378$, and $c=-0.008$. $\boxed{4}$

(d) The model was fitted with $\lambda=0$, so we need to log all values when forecasting.

$$\begin{aligned}\log\hat{y}_{T+1|T} &= -0.008+0.766\log(y_T)+0.622\log(y_{T-3})-0.476\log(y_{T-4})\\ &\qquad +0.378\log(y_{T-7})-0.290\log(y_{T-8})\\ &= -0.008+0.766\log(1.180)+0.622\log(1.180)-0.476\log(1.214)\\ &\qquad +0.378\log(1.207)-0.290\log(1.263)\\ &= 0.1328.\end{aligned}$$

The 95% PI on the log scale is $0.1328\pm1.96\sqrt{0.000912}=[0.074,0.192]$ $\boxed{4}$

Reversing the log, we get $e^{0.1328}=1.142$ and PI [1.08,1.21] $\boxed{1}$

(d) This is true.

For one mark, they could argue that the process will "blow up" if $\phi < 1$ or $\phi > 1$ $\boxed{1}$

For three marks, they should show that the root of $1 - \phi z = 0$ is $z = 1/\phi$. For $|z| > 1$, we require $|\phi| < 1$. $\boxed{3}$

For two marks, they could show that the mean and variance of y_t only converge when $|\phi| < 1$. $\boxed{2}$

[Total: 20 marks]

— END OF QUESTION 4 —

QUESTION 5

(a) Three knots have been selected at times 1990, 1996 and 2000. Then a dynamic regression model has been fitted for every possible choice of knot combination. $\boxed{1}$

The model with the lowest AIC value has been selected. $\boxed{1}$

The fitted model is

$$y_t = \beta_1 t + \beta_2(t-1990)_+ + \beta_3(t-1996)_+ + \beta_4(t-2000)_+ + n_t$$

$$(1-\phi B)(1-\Phi B^4)n_t = (1-\Theta B^4)\varepsilon_t$$

where $(u)_+ = max(u,0)$, $\beta_1 = 0.0005$, $\beta_2 = -0.0701$, $\beta_3 = 0.0806$, $\beta_4 = -0.0364$, $\phi = 0.5466$, $\Phi = 0.8546$ and $\Theta = -0.4459$. $\boxed{4}$

(b) This could be generalized and automated by having a list of all possible knot positions on a grid with a minimal knot separation. Then every combination of knots could be considered. $\boxed{2}$

One problem will be computation time. Every possible combination of a large number of knots will be a huge number of models. $\boxed{2}$

Another problem will be knot separation. Knots too close together will lead to degenerate models. $\boxed{2}$

(c) When producing forecasts using this model, we are assuming tha the future trend has no more breakpoints and will continue linearly from the end of the fitted trend. With an ETS model, although it also has a locally linear trend, we allow for it to change stochastically. An important consequence is that the width of the PIs will be much narrower for the dynamic regression model $\boxed{4}$

(d) A model with Fourier terms assumes the seasonal pattern is unchanging over time. Clearly with this series the seasonal pattern has changed quite a lot. Seasonal ARIMA models allow for changing seasonality by modelling seasonal differences. $\boxed{2}$

(e) While the point forecasts look ok, the prediction intervals are much too narrow. This is because of the assumptions about the trend. $\boxed{2}$

[Total: 20 marks]

— END OF QUESTION 5 —