

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

## **QUESTION 1**

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.

1. Linear regression models are simplistic because the real world is nonlinear.
2. MAE, RMSE, MAPE and MASE are all similar measures; hence it does not matter which one we use.
3. Always choose the model with the smallest sum of squared errors.
4. A good test of whether a model will produce good forecasts is that the residuals are stationary.
5. Forecasts should never give the same value for all forecast horizons.
6. Regression models are not useful for forecasting because we always need to provide forecasts of the predictors.

[5 marks each. Total 20 marks]

**Total: 0 marks**

— END OF QUESTION 1 —

## QUESTION 2

Figures 1–3 relate to the weekly mortality rate in Australia from 2015 – 2022. The mortality rate is defined as the number of deaths per thousand people in Australia in each week.

```
aus_mortality |>
  autoplot(rate) +
  labs(title = "Weekly Australian mortality rate", y = "Deaths per thousand people")
```

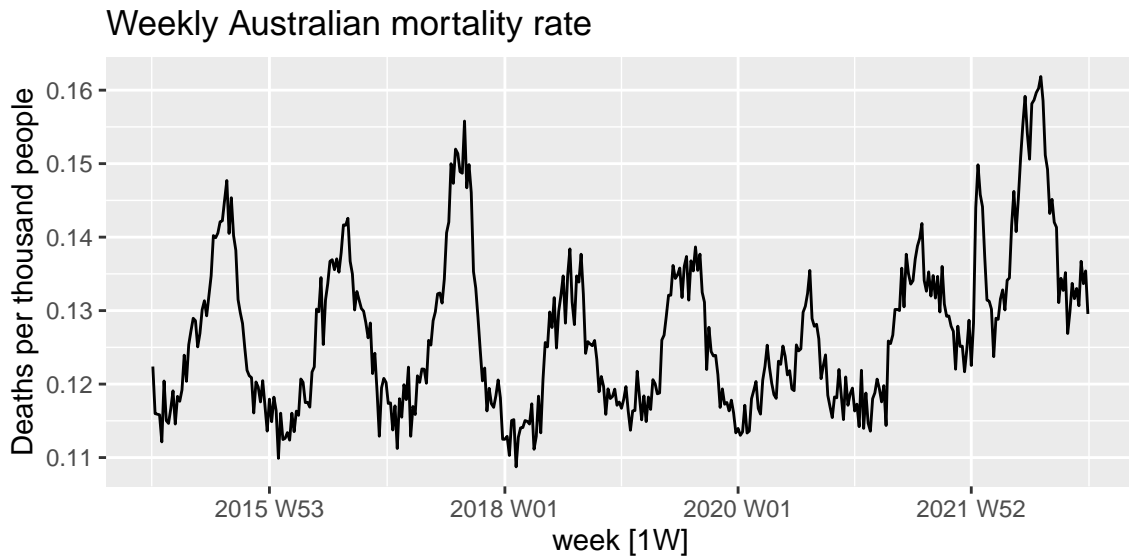


Figure 1:

```
aus_mortality |>
  gg_season(rate, labels = "right")
```

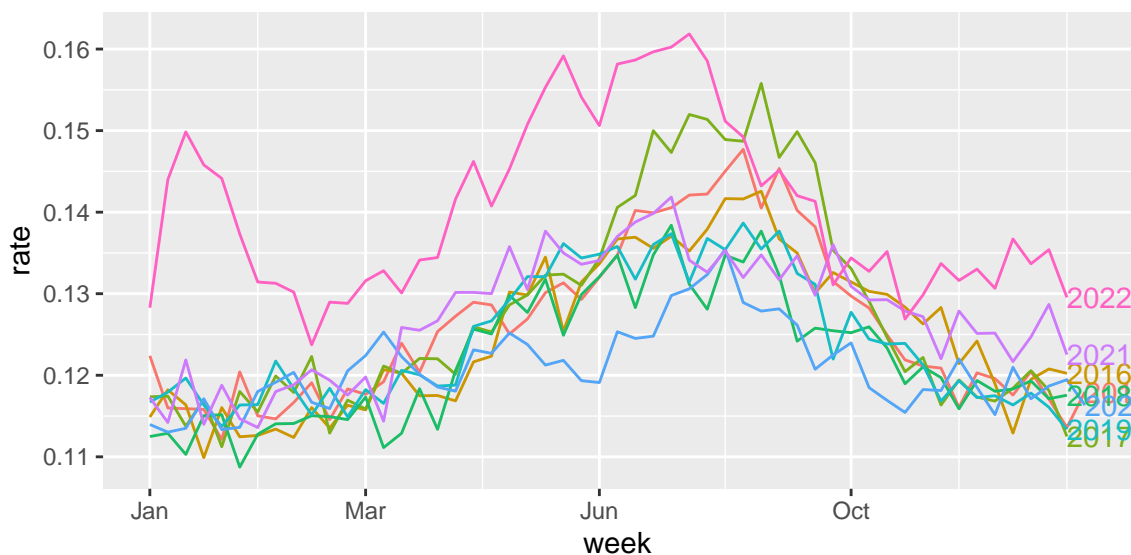


Figure 2:

```
aus_mortality |>
  model(stl = STL(rate ~ trend(window = 9) + season(window = "periodic"), robust = TRUE)) |>
```

```
components() |>
autoplot()
```

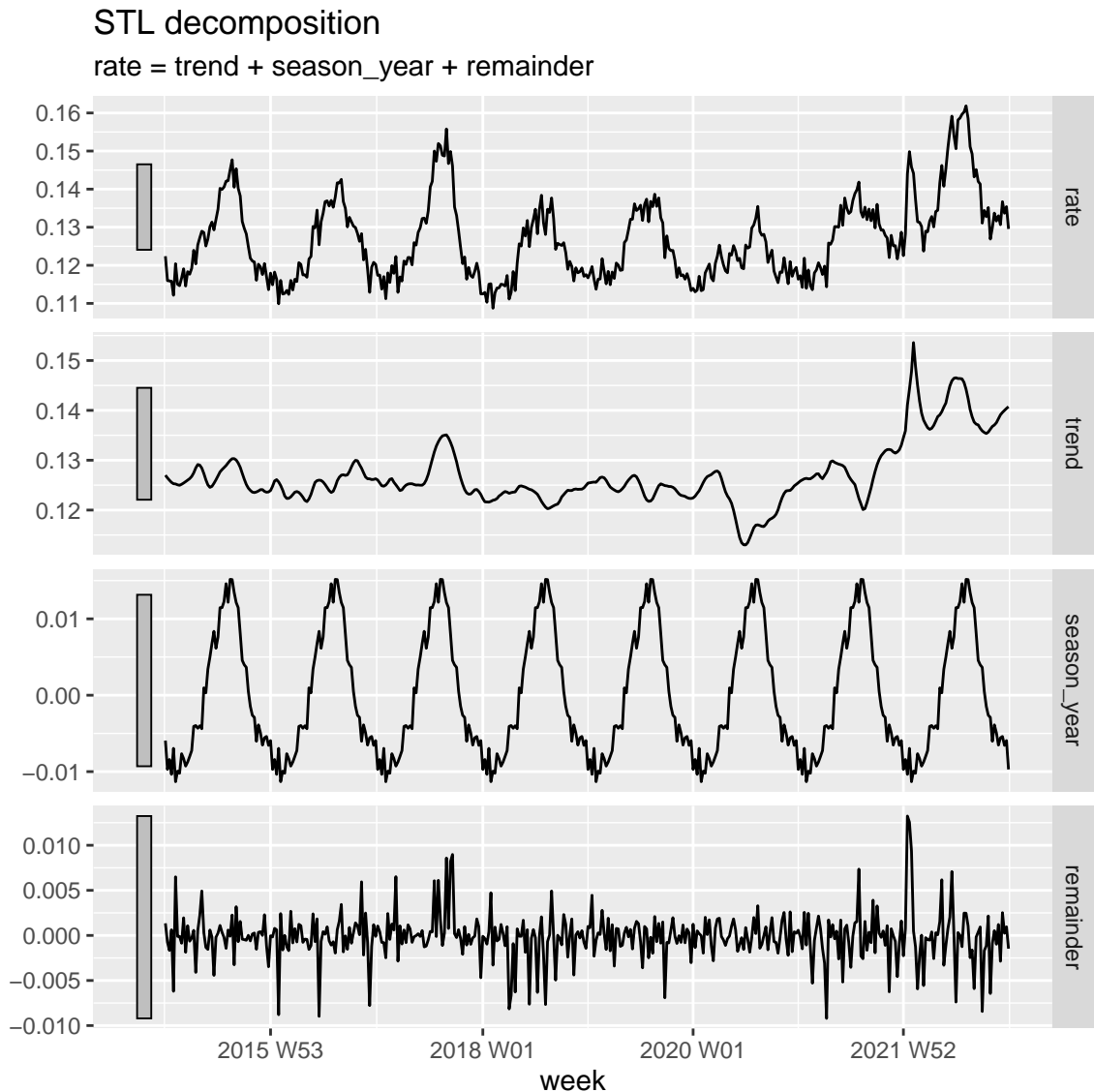


Figure 3:

1. Using Figures 1–3, describe the weekly mortality rate for Australia. Carefully comment on the interesting features of all three plots. [6 marks]
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]
3. You have been asked to provide forecasts for the next year for Australian weekly mortality rates. Consider applying each of the methods and models below. Comment, in a few words each, on whether each one is appropriate for forecasting the data. No marks will be given for simply guessing whether a method or a model is appropriate without justifying your choice. [10 marks]

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

- (a) Seasonal naïve method using annual seasonality.
- (b) Seasonal naïve method using weekly seasonality.

- (c) An STL decomposition with an ETS to forecast the seasonally adjusted component, and seasonal naïve for the seasonal component.
- (d) Holt's method with damped trend.
- (e) ETS(A,N,A).
- (f) ETS(M,A,M) with annual seasonality.
- (g) ARIMA(0,1,52).
- (h) ARIMA(2,0,2)(0,1,0)<sub>52</sub>.
- (i) Regression with Fourier terms for the annual seasonality.
- (j) Dynamic regression with Fourier terms for the annual seasonality.

Total: 0 marks

— END OF QUESTION 2 —

## QUESTION 3

1. Why can't you fit a seasonal ETS model to this data set? [3 marks]
2. You decide to use the STL decomposition shown in Figure 3, with ETS used for the seasonally adjusted data.

```
my_model <- decomposition_model(  
  STL(rate ~ season(window = "periodic")),  
  ETS(season_adjust ~ season("N")),  
  SNAIVE(season_year)  
)  
  
fit <- aus_mortality |>  
  model(my_model)  
  report(fit)
```

```
## Series: rate  
## Model: STL decomposition model  
## Combination: season_adjust + season_year  
##  
## =====  
##  
## Series: season_adjust  
## Model: ETS(M,N,N)  
## Smoothing parameters:  
##   alpha = 0.6008114  
##  
## Initial states:  
##   l[0]  
## 0.1273692  
##  
## sigma^2: 6e-04  
##  
##      AIC      AICc      BIC  
## -2273.114 -2273.056 -2261.007  
##  
## Series: season_year  
## Model: SNAIVE  
##  
## sigma^2: 0
```

Write down the equations for the fitted model, including the STL equation, and the equations for the components. [6 marks]

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

```
augment(fit) |>  
  autoplot(.innov)
```

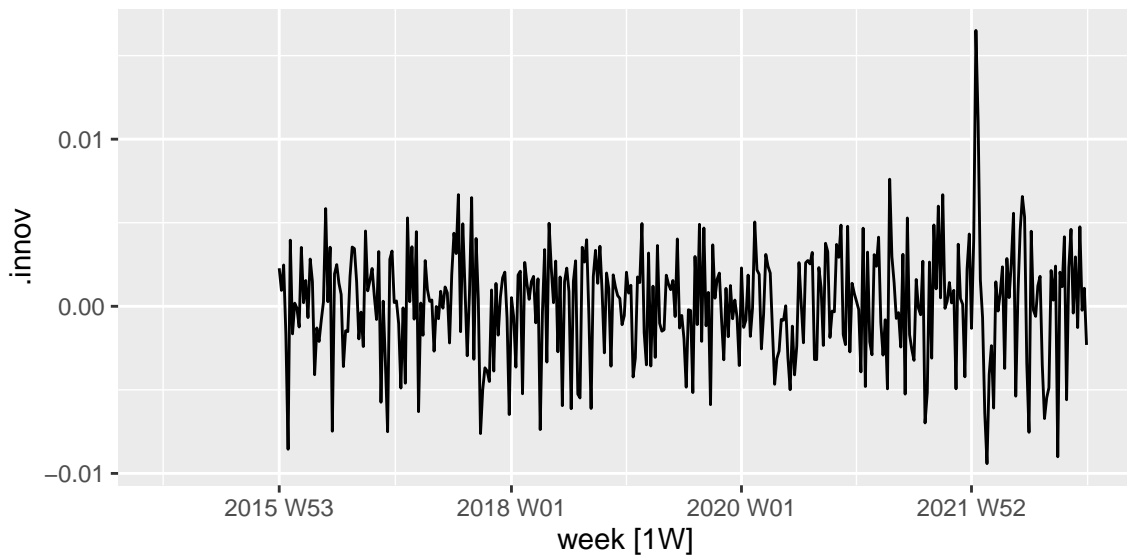


Figure 4:

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

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

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

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

```
fit |>
  forecast(h = "1 year") |>
  autoplot(aus_mortality)
```

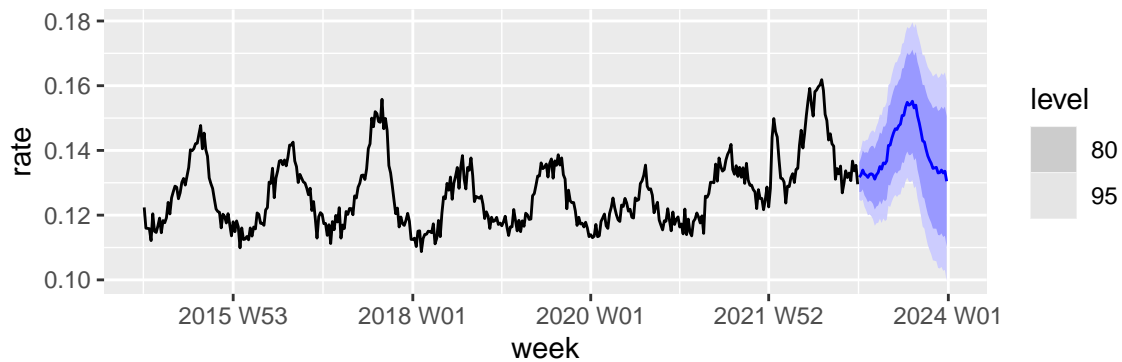


Figure 5:

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

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

Total: 0 marks

— END OF QUESTION 3 —

## QUESTION 4

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

```
fit <- aus_mortality |>
  model(ARIMA(rate ~ fourier(K = 2)))
report(fit)
```

```
## Series: rate
## Model: LM w/ ARIMA(2,1,1) errors
##
## Coefficients:
##          ar1      ar2      ma1  fourier(K = 2)C1_52  fourier(K = 2)S1_52
##          0.5361  0.2474 -0.9491                -5e-04                -0.0107
## s.e.      0.0564  0.0526  0.0270                 1e-03                 0.0010
##          fourier(K = 2)C2_52  fourier(K = 2)S2_52
##                               -0.0031                 1e-04
## s.e.                          0.0007                 7e-04
##
## sigma^2 estimated as 1.173e-05:  log likelihood=1778.69
## AIC=-3541.37  AICc=-3541.02  BIC=-3509.11
```

1. Write down the full model using backshift notation. [4 marks]
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]
3. Comment on the model diagnostics shown in Figure 6 and the output below. [5 marks]

```
fit |>
  gg_tsresiduals()
```

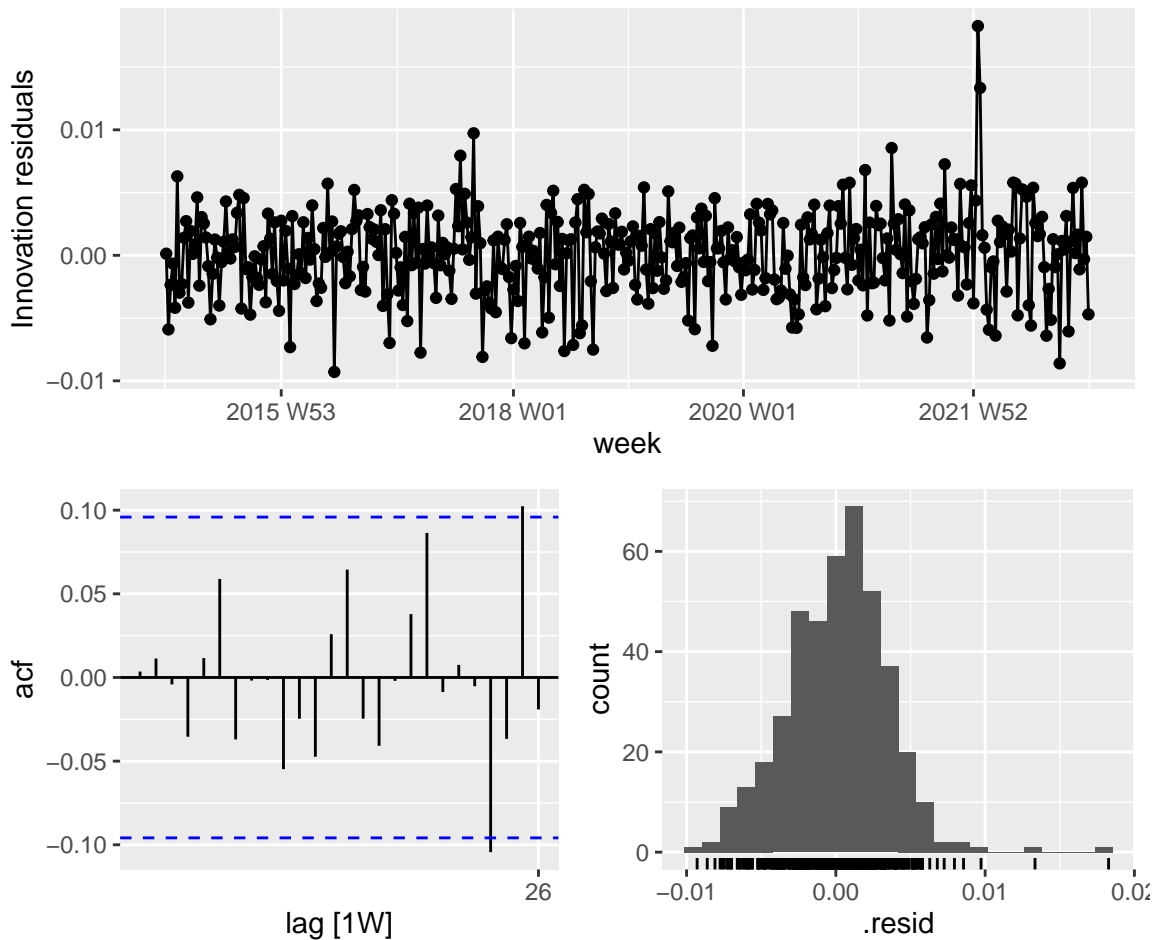


Figure 6:

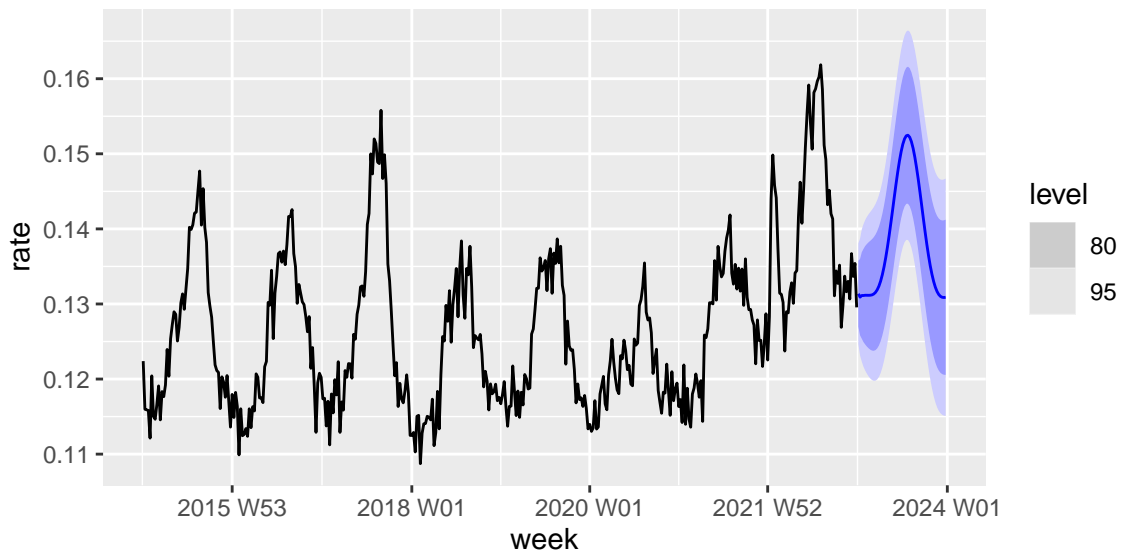
```
augment(fit) |>
  features(.innov, ljung_box, lag = 52, dof = 3)
```

```
## # A tibble: 1 x 3
##   .model                lb_stat lb_pvalue
##   <chr>                 <dbl>  <dbl>
## 1 ARIMA(rate ~ fourier(K = 2)) 37.1    0.894
```

4. Figure 7 shows forecasts with prediction intervals for the next year.

```
fit |>
  forecast(h = "1 year") |>
  autoplot(aus_mortality)
```





**Figure 7:**

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

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

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

**Total: 0 marks**

— END OF QUESTION 4 —

## QUESTION 5

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.

```
my_model <- decomposition_model(  
  STL(rate ~ season(window = "periodic")),  
  ETS(season_adjust ~ season("N")),  
  SNAIVE(season_year)  
)  
stretch <- aus_mortality |>  
  stretch_tsibble(.init = 105)  
fit <- stretch |>  
  model(  
    stl_ets = my_model,  
    dyn_regr = ARIMA(rate ~ fourier(K = 2)),  
    snaive = SNAIVE(rate)  
  )  
  
fc1 <- fit |> forecast(h = "6 weeks")  
fc1 |>  
  accuracy(aus_mortality,  
    measures = list(point_accuracy_measures, interval_accuracy_measures)  
  )
```

```
## # A tibble: 3 x 11  
##   .model .type      ME      RMSE      MAE      MPE      MAPE      MASE      RMSSE      ACF1      winkler  
##   <chr>  <chr>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>  
## 1 dyn_regr Test  0.000236 0.00542 0.00394 0.0808   3.04  0.613 0.623 0.689 0.0337  
## 2 snaive  Test  0.00261 0.00922 0.00685 1.73     5.18  1.07  1.06 0.798 0.0571  
## 3 stl_ets  Test  0.000186 0.00547 0.00405 0.0849   3.14  0.631 0.628 0.638 0.0347
```

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]
2. What do you conclude from the above output about the three models? [2 marks]
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]
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 dynamic regression model. Why would we prefer to select a model using MAE rather than ME? [3 marks]
5. The MASE and RMSSE accuracy measures use scaled errors. Why is scaling errors unnecessary in this analysis? [2 marks]
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]
7. **ETC3550 only:** The Winkler score is defined as

$$W_{\alpha,t} = \begin{cases} (u_{\alpha,t} - l_{\alpha,t}) + \frac{2}{\alpha}(l_{\alpha,t} - y_t) & \text{if } y_t < l_{\alpha,t} \\ (u_{\alpha,t} - l_{\alpha,t}) & \text{if } l_{\alpha,t} \leq y_t \leq u_{\alpha,t} \\ (u_{\alpha,t} - l_{\alpha,t}) + \frac{2}{\alpha}(y_t - u_{\alpha,t}) & \text{if } y_t > u_{\alpha,t} \end{cases}$$

where the  $100(1 - \alpha)\%$  prediction interval is given by  $[\ell_{\alpha,t}, u_{\alpha,t}]$ . 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]

**ETC5550 only:** Mortality rates vary with age and sex, so a better model would allow for different rates of change based on demographic characteristics. What would be the problem in developing a separate model for each age and sex group? [5 marks]

<b>Total: 0 marks</b>
-----------------------

— END OF QUESTION 5 —