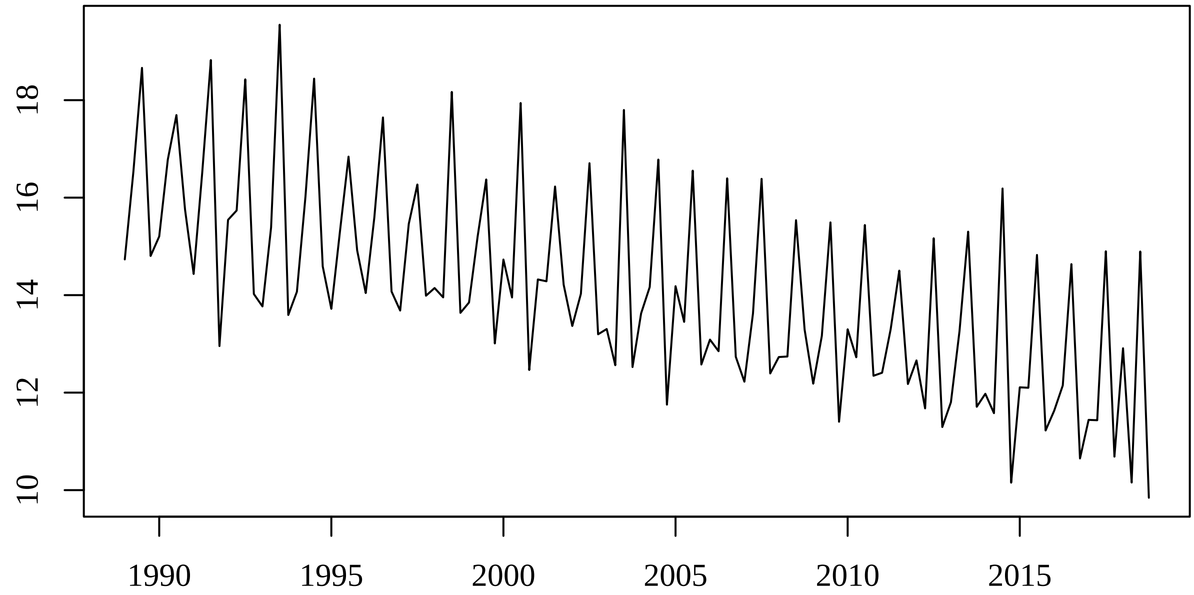
Question 1.

The following plot shows $n = 120$ observations on a quarterly time series from 1989q1 to 2018q4, to be used as the estimation sample. The four observations for 2019q1 to 2019q4 are reserved for forecast evaluation purposes.



(a) Describe the features evident in this plot that would be included in the specification of the deterministic trend equation for the time series.

A linear trend (downwards) is clear in the plot so a time trend would be included.

Also it appears there is a clear repeating seasonal pattern, so quarterly dummies would be included.

A model selection search was carried out for this time series. Two deterministic specifications were tried:

- intercept and linear trend (`NoQD`)
- intercept, linear trend and quarterly dummy variables (`QD`)

in combination with ARMA(p,q) models with a variety of p and q. The values of the AICc and Ljung-Box test p-values are tabulated below.

p	q	AICc(NoQD)	LBp(NoQD)	AICc(QD)	LBp(QD)
0	0	462.42	0.000	247.66	0.000
1	0	448.32	0.000	213.64	0.027
2	0	420.09	0.000	208.60	0.065
3	0	258.62	0.000	207.11	0.042
4	0	260.54	0.000	204.76	0.073
5	0	261.23	0.000	206.81	0.064
6	0	263.54	0.000	209.22	0.064
7	0	227.57	0.110	203.39	0.485
8	0	228.99	0.087	204.92	0.467
0	1	373.16	0.000	192.37	0.823
1	1	373.87	0.000	194.12	0.636
2	1	348.77	0.000	193.81	0.517
0	2	368.15	0.000	194.26	0.626
1	2	362.35	0.000	196.89	0.191
2	2	342.13	0.000	195.94	0.311

(b) What is the best model *without* quarterly dummmies? Justify your choice.

The AR(7) model is the best model without quarterly dummies. It has the smallest AICc amongst these models, and passes the residual autocorrelation test (LBp>0.05).

(c) What is the best model amongst all of those considered? Justify your choice.

The MA(1) model with quarterly dummies is the best model overall. It has the smallest AICc and passes the residual autocorrelation test (LBp>0.05).

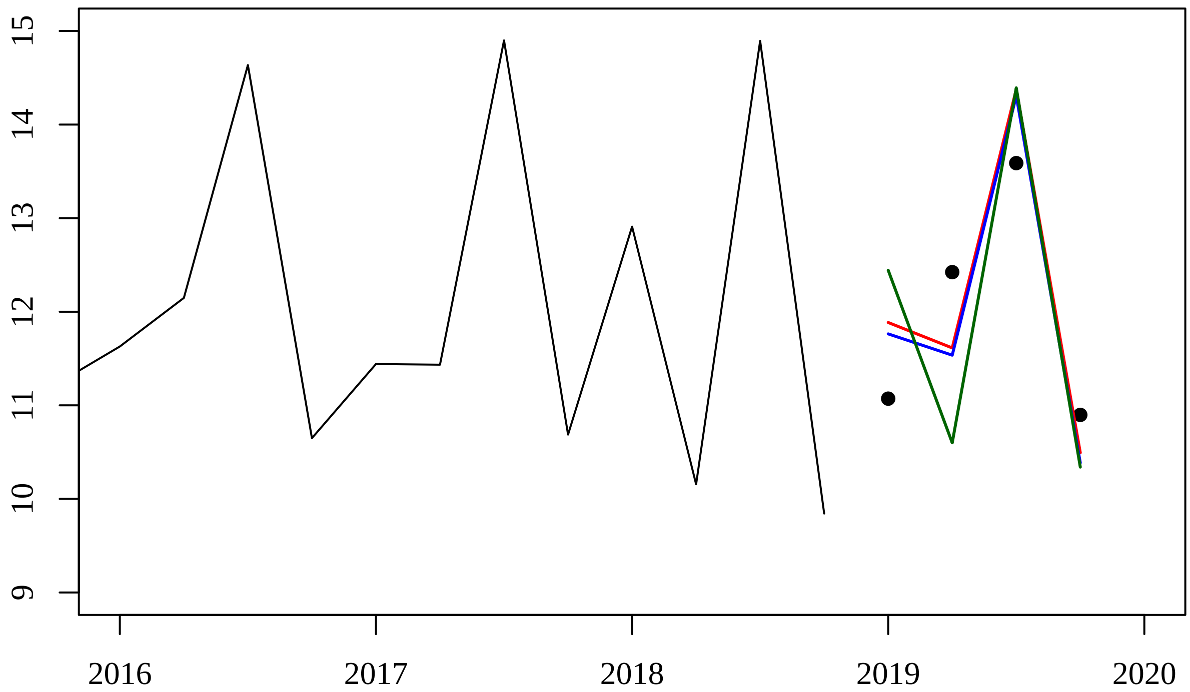
Forecasts and RMSE for the four quarters of 2019 were computed from three models

A: ARMA(0,1) with quarterly dummies, RMSE= 0.725

B: ARMA(7,0) with quarterly dummies, RMSE= 0.716

C: ARMA(7,0) *without* quarterly dummies, RMSE= 1.242

and plotted below, along with the actual values for this period.



(d) Discuss the accuracy of the forecasts from these models, and also how they relate to the in-sample AICc values.

Model B has the smallest RMSE is therefore considered most accurate.

Model A has slightly worse RMSE, but the plot makes clear that there is very little practical difference between the Model A and B forecasts.

The Model C forecasts have substantially worse accuracy than the other models, most obviously for the first two observations of the forecast period.

The AICc suggest that Model A is preferred to Model B in-sample, but this ordering is reversed in the forecast period. Model C has largest AICc of the three models, and this in-sample inferiority carries over to the forecast period.

The coefficient estimates for Model A are as follows:

```
   ma1 intercept      time        Q1        Q2        Q3
-0.855    14.980    -4.487     0.552     1.048     3.846
```

(e) Write out the estimated model in equation form.

$$Y_t = 14.980 - 4.487\,\mathrm{Time}_t + 0.552\,Q_{1,t} + 1.048\,Q_{2,t} + 3.846\,Q_{3,t} + Z_t$$
$$Z_t = U_t - 0.855U_{t-1}$$

(f) According to Model A, in which quarter does Y_t have its highest unconditional mean value? In which quarter does Y_t have its lowest unconditional mean value?

Highest: quarter 3, since it has the largest coefficient.

Lowest: quarter 4, since the q1,2,3 all have positive coefficients.

(g) On the basis of the information shown so far, give an explanation for why Models A and B can give such similar forecasts despite their quite different functional forms.

Model A has an invertible MA component, which could therefore be approximated by an AR model with lag order "long enough" to to produce a good approximation. The MA(1) parameter estimate in Model A is quite large (0.855 is not far from the unit circle) so it is not surprising the 7 AR lags are required to produce a close approximation, and hence similar forecasts.

(h) When we construct a 95% prediction interval, what does the probability / proportion 95% refer to?

In repeated samples from the population distribution of the forecast period, each of the prediction intervals (i.e. at each h) should include / "cover" the actual value of the forecast series in 95% of the samples.

The following table shows the normal 95% prediction intervals for models B and C, and their width (respectively `B:Lower`, `B:Upper`, `B:Width` for model B, and `C:Lower`, `C:Upper`, `C:Width` for model C). Column `Actual` shows the actual values of Y_t in the forecast period.

	Actual	B:Lower	B:Upper	B:Width	C:Lower	C:Upper	C:Width
h=1	11.071	10.74	12.79	2.05	11.32	13.57	2.25
h=2	12.423	10.26	12.81	2.56	9.10	12.10	3.00
h=3	13.588	13.03	15.59	2.56	12.89	15.90	3.01
h=4	10.898	9.10	11.67	2.57	8.83	11.85	3.01

(i) Compare the quality of the prediction intervals from models B and C on the basis of their inclusion / coverage of the actual value of Y_t.

The model B prediction intervals include the actual value of Y_t for every h.

The model C prediction intervals do not include the actual value of Y_t for $h = 1, 2$, but they do for $h = 3, 4$.

The model B prediction intervals are superior on this criterion.

(j) Compare the quality of the prediction intervals from models B and C on the basis of their width.

The model B prediction intervals are substantially narrower (roughly 10-15%) compared to those from model C, so the model B prediction intervals are superior in this respect as well.

(k) Compare the overall forecast accuracy and prediction interval quality of Models B and C. To what do you attribute the difference in performance between the two models?

The model B point forecasts are more accurate according the RMSE, and the model B prediction intervals are superior for both coverage and width.

Models B and C differ only in the inclusion of quarterly dummies in Model B, so clearly getting that choice correct for this time series was critical for overall forecasting performance.