Econ 493 B1 - Winter 2023

Homework 2 - Solution

Exercise 1

You work for the International Monetary Fund in Washington DC, monitoring Singapore's real consumption expenditures. Using a sample of real consumption data (measured in billions of 2005 Singapore dollars), y_t, t = 1990Q1, …2016Q4, you estimate the linear consumption trend model

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

obtaining the estimates $\hat{\beta}_0 = 0.51$, $\hat{\beta}_1 = 2.30$, and $\hat{\sigma}_e = 4$.

a. Based upon your estimated trend model, construct a point forecast for 2018Q4.

1990Q1: $t = 1$
2016Q4: $t = 108$
2018Q4: $t = 116$

Then, the point forecast is $\hat{y}_{2018Q4|2016Q4} = 0.51 + 2.30 \times 116 = 267.31$.

b. Based upon your estimated trend model, construct an interval forecast for 2018Q4.

Then, $267.31 \pm 1.96 \times \hat{\sigma}_e = 267.31 \pm 1.96 \times 4$ and the 95% prediction interval is (259.47,275.15).

Exercise 2

Describe how you would construct a purely seasonal model for the following monthly series. In particular, what dummy variable(s) would you use to capture the relevant effects?

a. A sporting goods store finds that detrended monthly sales are roughly the same for each month in a given three-month season. For example, sales are similar in the winter months of January, February and March, in the spring months of April, May and June, and so on.

Four **monthly** dummies indicating the quarter would do the job:

$D_1 = \{1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, \ldots\}$
$D_2 = \{0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, \ldots\}$
$D_3 = \{0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, \ldots\}$
$D_4 = \{0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, \ldots\}$

b. A campus bookstore finds that detrended sales are roughly the same for all first, all second, and all third months of each trimester. For example, sales are similar in January, April, July, and October, the first months of the first, second, third, and fourth trimesters, respectively.

Three monthly dummies indicating the month within the trimester:

$D_1 = \{1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0 \ldots\}$
$D_2 = \{0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0 \ldots\}$
$D_3 = \{0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1 \ldots\}$

c. A Christmas ornament store is only open in November and December, so sales are zero in all other months.

This is a rather complicated situation. One could use two dummies, one for November-December, and one for other (or three, one for November, one for December, and one for other), perhaps even imposing the constraint that the coefficient on the "other'' dummy is zero.

Exercise 3 (R)

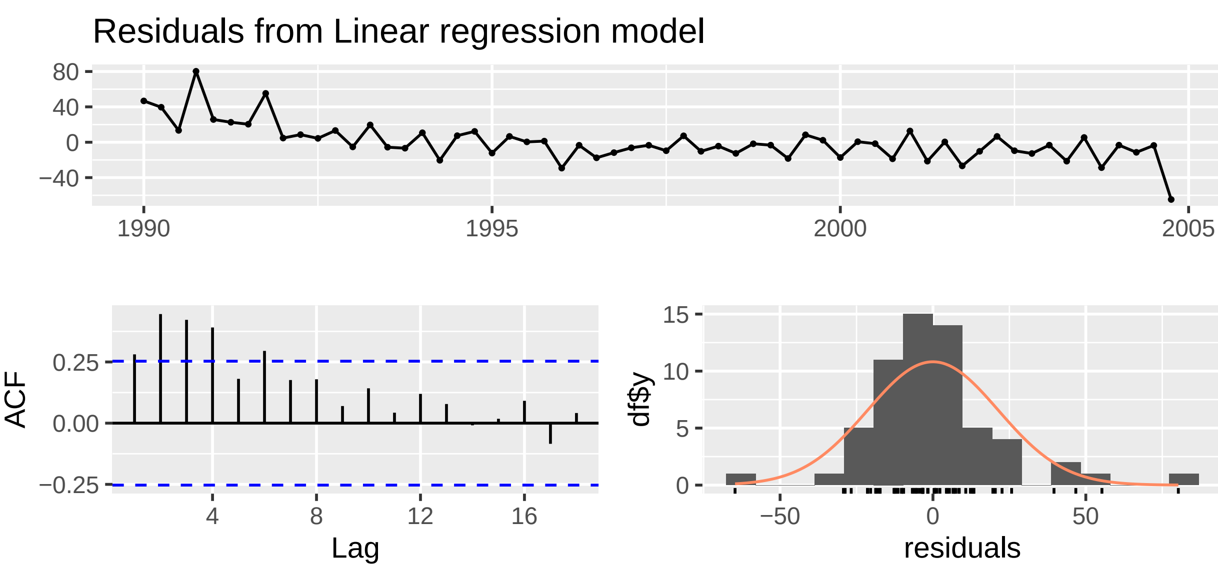
Consider the data set of quarterly Australian beer production data from 1990.

a. Estimate a linear model with seasonal dummies as predictors using data from 1990Q1 to 2004Q4. Evaluate the residuals. Compute the AIC and BIC.

```
beer <- window(ausbeer, start=1990, end=c(2004,4))
# seasonal dummies
fit.beer1 <- tslm(beer ~ season)
summary(fit.beer1)
```

```
##
## Call:
## tslm(formula = beer ~ season)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -64.73 -11.48  -3.27   7.65  80.27
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   438.27       5.67   77.29  < 2e-16 ***
## season2       -36.87       8.02   -4.60  2.5e-05 ***
## season3       -22.67       8.02   -2.83   0.0065 **
## season4        80.47       8.02   10.03  4.0e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22 on 56 degrees of freedom
## Multiple R-squared:  0.821,  Adjusted R-squared:  0.811
## F-statistic: 85.4 on 3 and 56 DF,  p-value: <2e-16
```

```
checkresiduals(fit.beer1)
```




```
##
```

```
##  Breusch-Godfrey test for serial correlation of order up to 8
##
## data:  Residuals from Linear regression model
## LM test = 25, df = 8, p-value = 0.0016
```

The residuals exhibit some downward linear trend and strong autocorrelation.

```r
CV(fit.beer1)
```

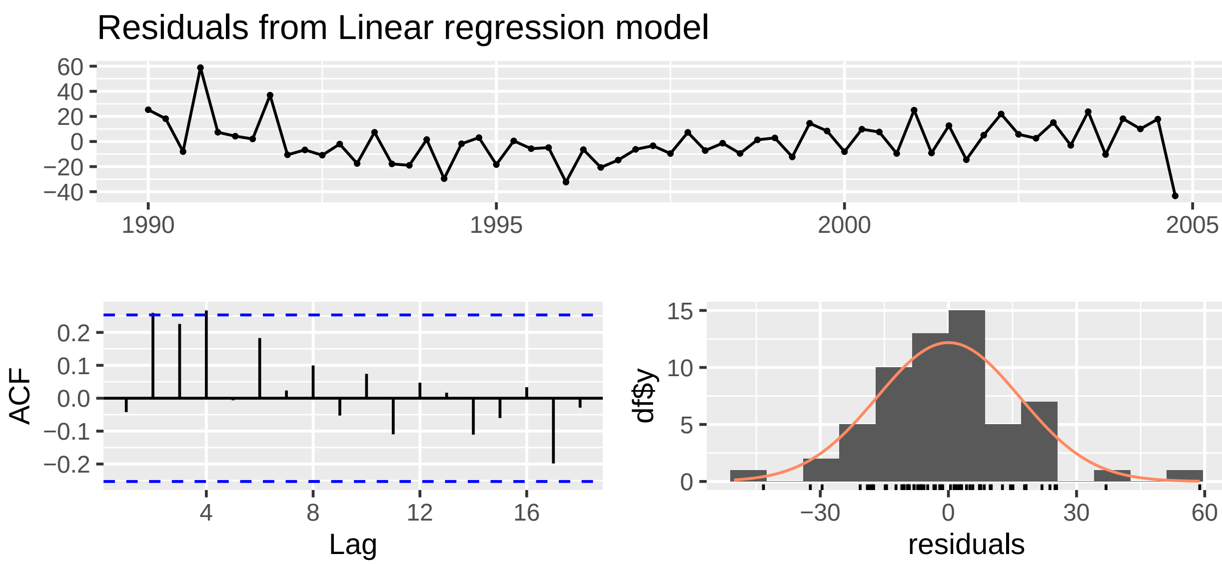
```
##        CV       AIC      AICc       BIC     AdjR2
## 516.79337 376.57944 377.69055 387.05117   0.81102
```

b. Estimate a linear model with a trend and seasonal dummies as predictors using data from 1990Q1 to 2004Q4. Evaluate the residuals. Compute the AIC and BIC.

```r
beer <- window(ausbeer, start=1990, end=c(2004,4))
# trend + seasonal dummies
fit.beer2 <- tslm(beer ~ trend + season)
summary(fit.beer2)
```

```
##
## Call:
## tslm(formula = beer ~ trend + season)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -43.31  -9.56  -1.56   7.78  58.84
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  460.457      5.841   78.83  < 2e-16 ***
## trend         -0.765      0.129   -5.91  2.2e-07 ***
## season2      -36.101      6.330   -5.70  4.8e-07 ***
## season3      -21.136      6.333   -3.34   0.0015 **
## season4       82.762      6.340   13.05  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.3 on 55 degrees of freedom
## Multiple R-squared:  0.89,   Adjusted R-squared:  0.882
## F-statistic:  112 on 4 and 55 DF,  p-value: <2e-16
```

```r
checkresiduals(fit.beer2)
```




```
##
##  Breusch-Godfrey test for serial correlation of order up to 8
##
## data:  Residuals from Linear regression model
## LM test = 16.7, df = 8, p-value = 0.033
```

The trend component improves the fit by capturing the downward trend in the data. The residuals still exhibit some autocorrelation.

```
CV(fit.beer2)
```

```
##        CV       AIC      AICc       BIC     AdjR2
## 335.68619 349.07509 350.66000 361.64116   0.88232
```

c. Which model is preferred? Explain.

The model with trend fits the data much better and is preferred based on both the AIC and BIC.

d. Evaluate the predictive performance of these models in the test set 2005Q1 to 2009Q4. Which model performs better?

```
# data and set up
beer1 <- window(ausbeer, start=1990,end=c(2009,4))
n.end <- 2004.75 # 2004Q4
# set matrix for storage, 20 obs in test set
pred <- matrix(rep(NA,80),20,4)
# loop
for(i in 1:20){
  pred[i,1] <- window(beer1,n.end+i*.25,n.end+i*.25) # actual
  tmp <- window(beer1,1990,n.end+(i-1)*.25)
  # compute forecasts 1
  fit.beer1 <- tslm(tmp ~ season)
  fcast1 <- forecast(fit.beer1, h=1)
  pred[i,2] <- fcast1$mean
  # compute forecasts 2
  fit.beer2 <- tslm(tmp ~ trend + season)
```

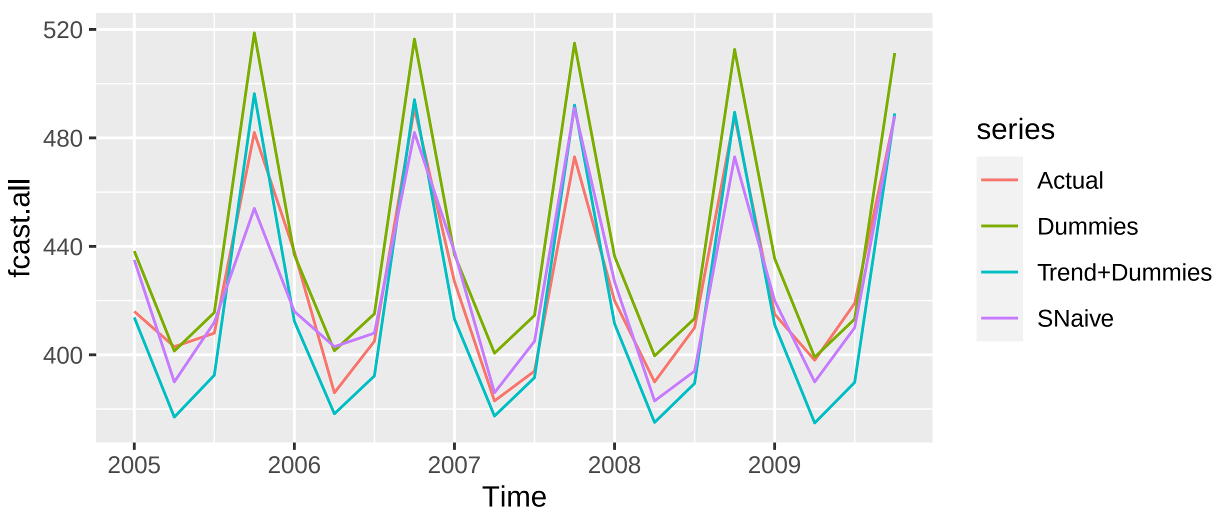
```r
  fcast2 <- forecast(fit.beer2, h=1)
  pred[i,3] <- fcast2$mean
  # compute forecasts 2
  fcast3 <- snaive(tmp, h=1)
  pred[i,4] <- fcast3$mean
}

# compute rmse
beerfit1 <- pred[,1] - pred[,2]
beerfit2 <- pred[,1] - pred[,3]
beerfit3 <- pred[,1] - pred[,4]
rmse1 <- sqrt(mean(beerfit1^2, na.rm=TRUE))
rmse2 <- sqrt(mean(beerfit2^2, na.rm=TRUE))
rmse3 <- sqrt(mean(beerfit3^2, na.rm=TRUE))

# display rmse
cbind(rmse1,rmse2,rmse3)
```

```
##       rmse1  rmse2  rmse3
## [1,] 19.317 15.378 13.261
```

```r
# plot
fcast.all <- ts(pred, start=2005, frequency=4, names=c("Actual","Dummies","Trend+Dummies
autoplot(fcast.all)
```



The model with trends performs substantially better in the test set than the model with seasonal dummies only. Overall, however, the seasonal naive model performs best. Why?

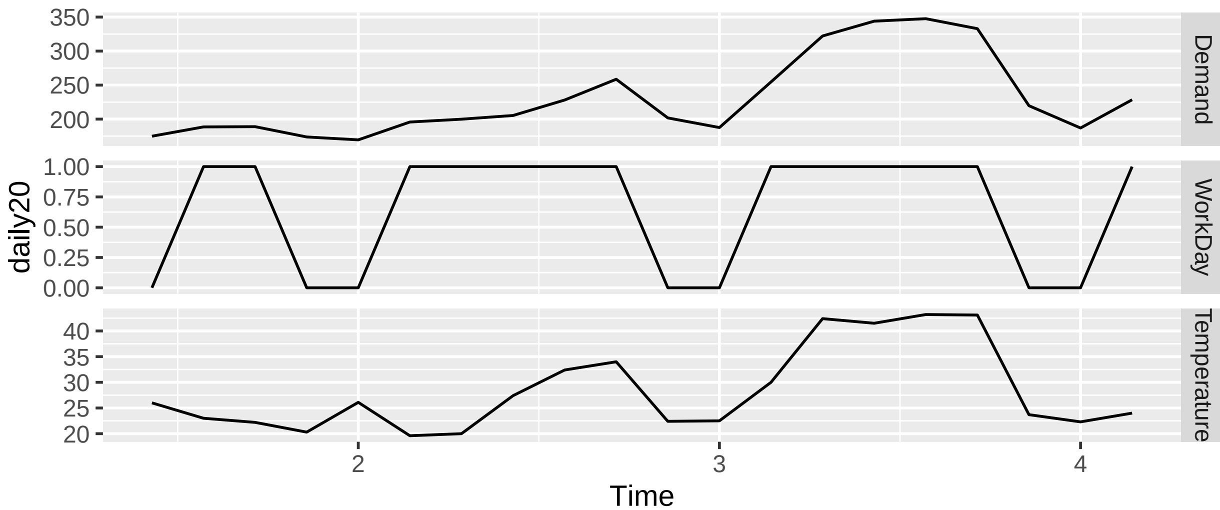
Exercise 4 (R)

Daily electricity demand for Victoria, Australia, during 2014 is contained in `elecdaily`. The data for the first 20 days can be obtained as follows.

```
daily20 <- head(elecdaily,20)
```

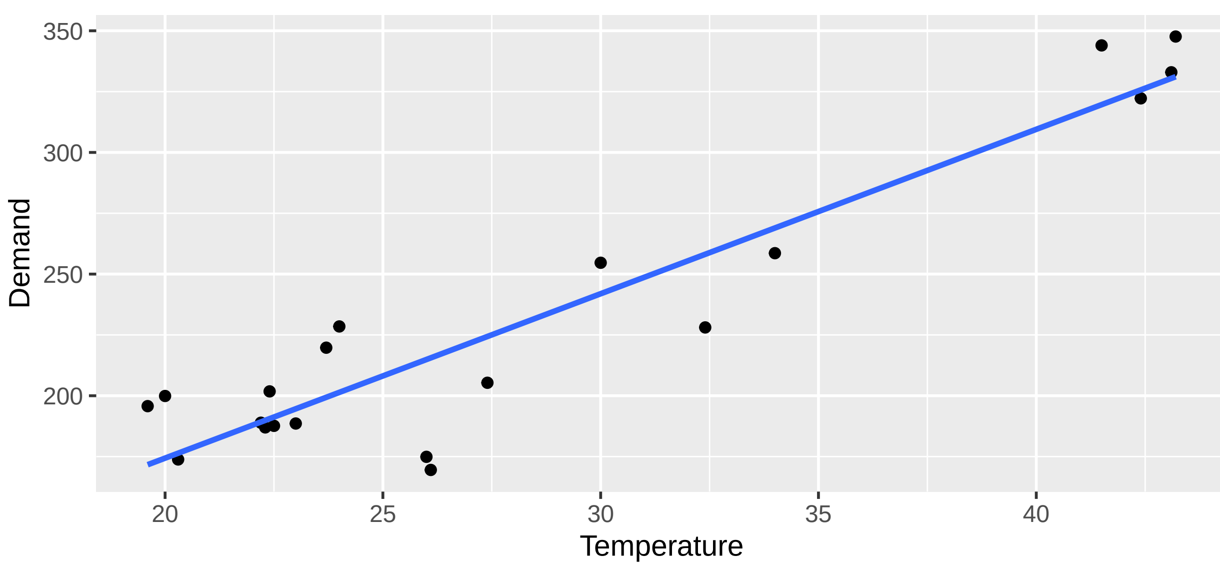
a. Plot the data and find the regression model for Demand with temperature as an explanatory variable. Why is there a positive relationship?

```
# plots
autoplot(daily20, facets=TRUE)
```




```
ggplot(aes(x = Temperature, y = Demand), data=as.data.frame(daily20)) +
    geom_point() +
    geom_smooth(method = "lm", se = FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
# linear regression
fit <- tslm(Demand ~ Temperature, data = daily20)
summary(fit)
```

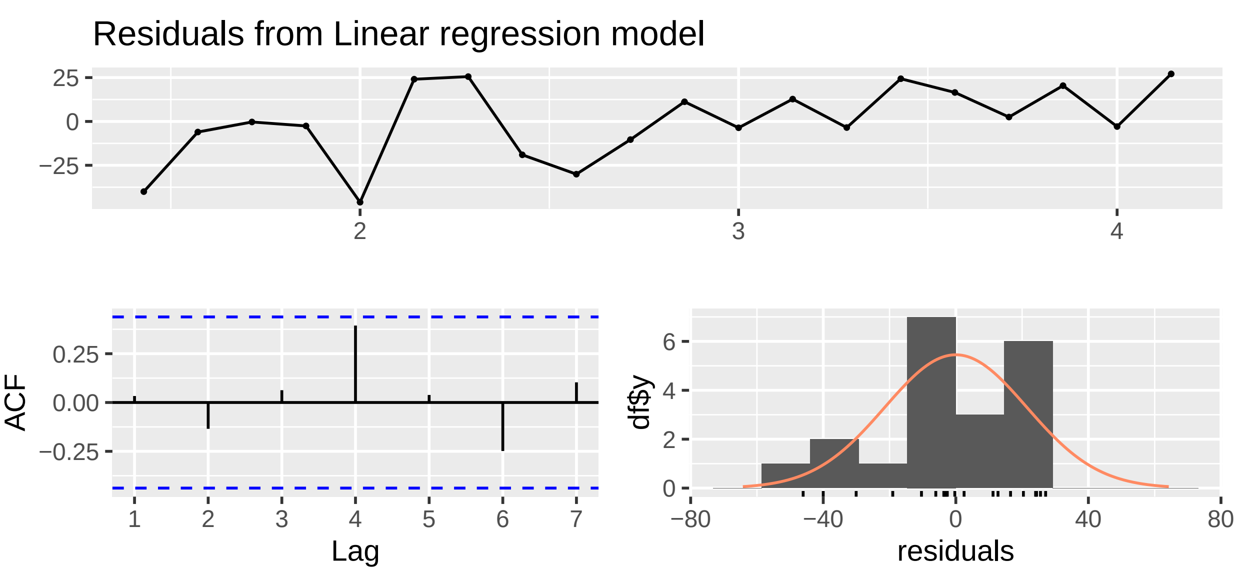
```
##
## Call:
## tslm(formula = Demand ~ Temperature, data = daily20)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -46.06  -7.12  -1.44  17.48  27.10
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   39.212     17.992    2.18    0.043 *
## Temperature    6.757      0.611   11.05  1.9e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22 on 18 degrees of freedom
## Multiple R-squared:  0.872,  Adjusted R-squared:  0.864
## F-statistic:  122 on 1 and 18 DF,  p-value: 1.88e-09
```

There is a positive relationship between temperature and electricity consumption.

Given the time of year, and the recorded temperature values, it is likely that electricity is being used for air conditioners. Since higher temperatures mean a higher demand for cooling, this leads to a positive relationship between temperature and electricity consumption.

b. Produce a residual plot. Is the model adequate? Are there any outliers or influential observations?

```
checkresiduals(fit)
```




```
##
##  Breusch-Godfrey test for serial correlation of order up to 5
##
## data:  Residuals from Linear regression model
## LM test = 3.81, df = 5, p-value = 0.58
```

Although the ACF tests are passed, there is a linear trend in the residuals. So the model looks inadequate.

c. Use the model to forecast the electricity demand that you would expect for the next day if the maximum temperature was 15° and compare it with the forecast if the with maximum temperature was 35°. Do you believe these forecasts?

```
(fc.low <- forecast(fit, newdata=data.frame("Temperature"=15)))
```

```
##        Point Forecast  Lo 80  Hi 80  Lo 95  Hi 95
## 4.2857         140.57 108.68 172.46 90.212 190.93
```

```
(fc.high <- forecast(fit, newdata=data.frame("Temperature"=35)))
```

```
##        Point Forecast  Lo 80 Hi 80  Lo 95  Hi 95
## 4.2857         275.71 245.23 306.2 227.57 323.86
```

The prediction for 35° looks reasonable, but the one for 15° assumes the trend continues to decrease for temperature values lower than 20, which is unlikely. Heating will mean it will increase for lower temperatures.

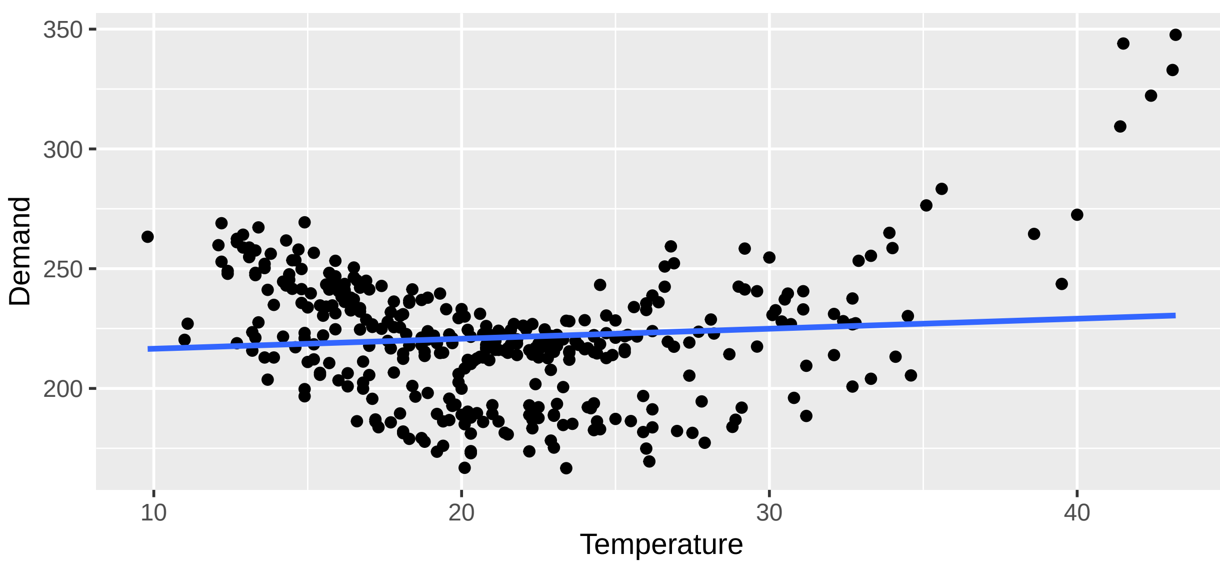
d. Give prediction intervals for your forecasts.

See above.

e. Plot Demand vs Temperature for all of the available data in `elecdaily`. What does this say about your model?

```
ggplot(aes(x = Temperature, y = Demand), data=as.data.frame(elecdaily)) +
    geom_point() +
    geom_smooth(method = "lm", se = FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



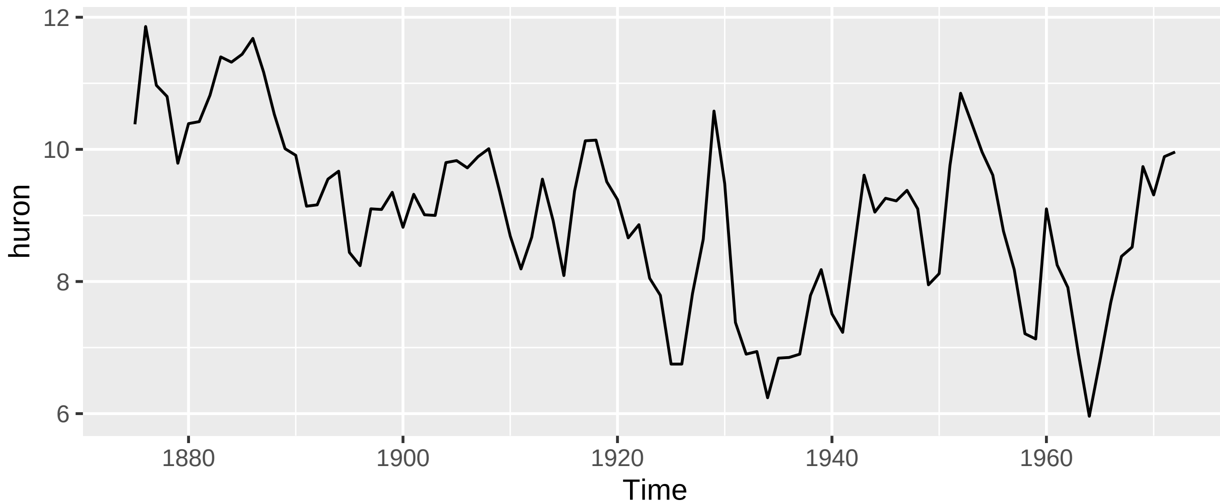
This shows the non-linear relationship clearly. Even limiting the data to above 20, there is a nonlinear relationship between demand and temperature. The model is inadequate.

Exercise 5 (R)

Data set `huron` gives the level of Lake Huron in feet from 1875-1972.

a. Plot the data and comment on its features.

```
autoplot(huron)
```



It seems that the water level was going down until around 1915 and then seems to have stabilised indicating a non-linear trend.

b. Fit a linear regression and compare this to a piecewise linear trend model with a knot at 1915.

```
# linear trend
fit.lin <- tslm(huron ~ trend)
summary(fit.lin)
```

```
##
## Call:
## tslm(formula = huron ~ trend)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5100 -0.7273  0.0008  0.7440  2.5357
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.20204    0.23011    44.3  < 2e-16 ***
## trend       -0.02420    0.00404    -6.0  3.5e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.13 on 96 degrees of freedom
## Multiple R-squared:  0.272,  Adjusted R-squared:  0.265
## F-statistic:   36 on 1 and 96 DF,  p-value: 3.55e-08
```

```
CV(fit.lin)
```

```
##        CV       AIC      AICc       BIC     AdjR2
##   1.30589 27.98370 28.23902 35.73860  0.26489
```

```
# piecewise linear trend
t <- time(huron)
tb <- ts(pmax(t-1915, 0))
fit.pw <- tslm(huron ~ t + tb)
summary(fit.pw)
```

```
##
## Call:
## tslm(formula = huron ~ t + tb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4963 -0.6624 -0.0714  0.8516  2.3922
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 132.9087    19.9769    6.65  1.8e-09 ***
## t            -0.0650     0.0105   -6.18  1.6e-08 ***
## tb            0.0649     0.0156    4.15  7.3e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.05 on 95 degrees of freedom
## Multiple R-squared:  0.384,  Adjusted R-squared:  0.371
## F-statistic: 29.6 on 2 and 95 DF,  p-value: 1e-10
```

```
CV(fit.pw)
```

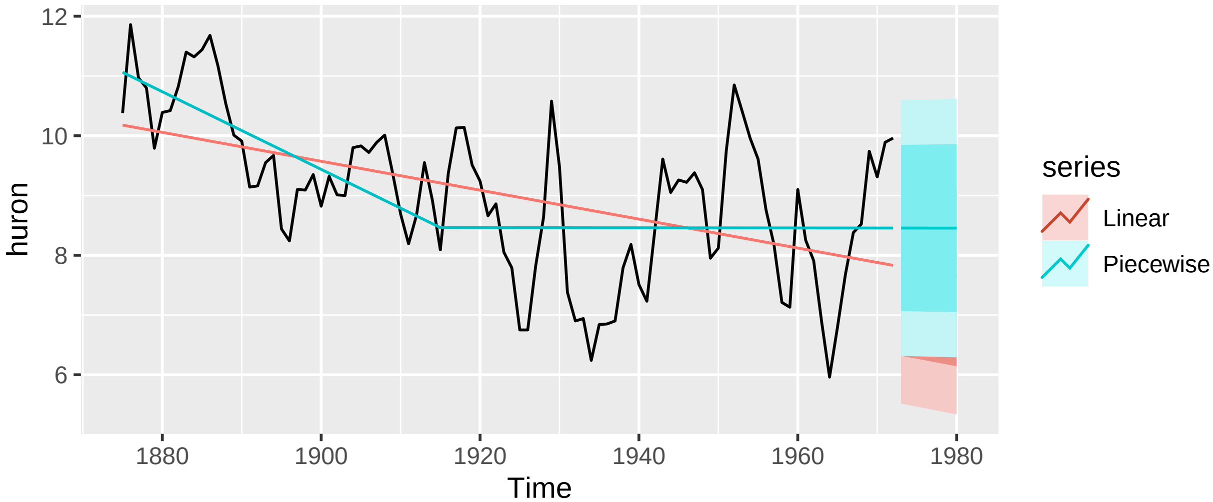
```
##        CV       AIC      AICc       BIC     AdjR2
##   1.12377 13.65947 14.08958 23.99934  0.37114
```

The two slope parameters almost add to zero, indicating the trend after 1915 is approximately flat.

c. Generate forecasts from these two models for the period upto 1980 and comment on these.

```
h <- 8
# forecast linear trend
fcasts.lin <- forecast(fit.lin, h=h)
# forecast piecewise linear trend
t.new <- t[length(t)] + seq(h)
tb.new <- tb[length(tb)] + seq(h)
newdata <- data.frame("t"=t.new,"tb"=tb.new)
fcasts.pw <- forecast(fit.pw, newdata = newdata)
# plot
autoplot(huron) +
  autolayer(fitted(fit.lin), series = "Linear") +
  autolayer(fitted(fit.pw), series = "Piecewise") +
```

```
  autolayer(fcasts.lin, series = "Linear") +
  autolayer(fcasts.pw, series="Piecewise")
```



The break in the trend in around 1915 seems reasonable. The projections from the piecewise linear trend show the water levels stabilise in contrast to the linear trend which shows a decline.

d. Repeat b. and c. with a knot at 1920 and comment on any differences.

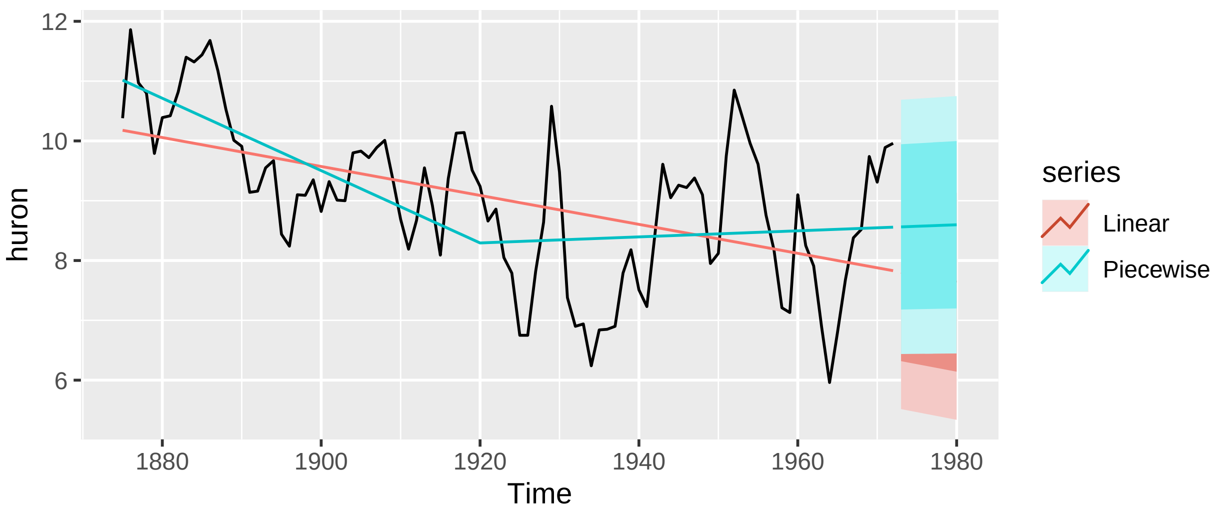
```
# piecewise linear trend
t <- time(huron)
tb <- ts(pmax(t-1920, 0))
fit.pw <- tslm(huron ~ t + tb)
summary(fit.pw)
```

```
##
## Call:
## tslm(formula = huron ~ t + tb)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.557 -0.626 -0.138  0.834  2.394
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 124.42259   17.19863    7.23  1.2e-10 ***
## t            -0.06048    0.00904   -6.69  1.5e-09 ***
## tb            0.06555    0.01490    4.40  2.8e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.04 on 95 degrees of freedom
## Multiple R-squared:  0.396,  Adjusted R-squared:  0.383
## F-statistic: 31.1 on 2 and 95 DF,  p-value: 4.12e-11
```

```
CV(fit.pw)
```

```
##        CV       AIC      AICc       BIC     AdjR2
##   1.10579 11.82093 12.25104 22.16080  0.38283
```

```
h <- 8
# forecast piecewise linear trend
t.new <- t[length(t)] + seq(h)
tb.new <- tb[length(tb)] + seq(h)
newdata <- data.frame("t"=t.new,"tb"=tb.new)
fcasts.pw <- forecast(fit.pw, newdata = newdata)
# plot
autoplot(huron) +
  autolayer(fitted(fit.lin), series = "Linear") +
  autolayer(fitted(fit.pw), series = "Piecewise") +
  autolayer(fcasts.lin, series = "Linear") +
  autolayer(fcasts.pw, series="Piecewise")
```



We need to be careful as the projections as the break point is subjectively chosen. 5 years later and we get a different projection with increasing levels.