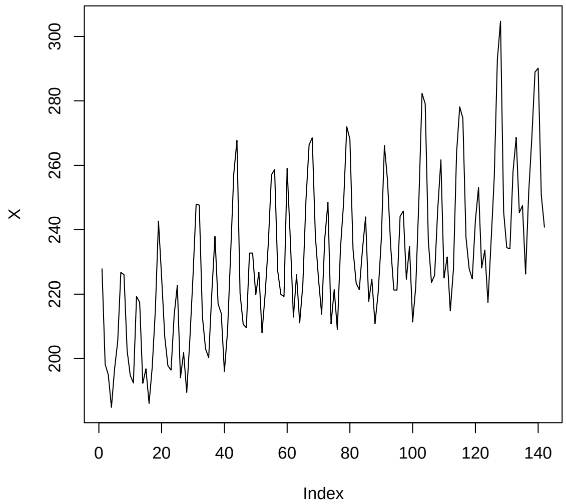
Exercises 2 Question 10 Solution

I used R to answer this question, but most statistical software packages will allow the same analysis.

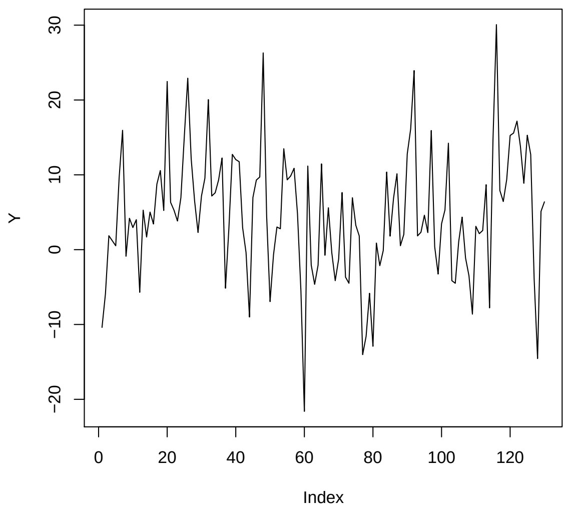
The data was saved to a file `KWhrs.dat`, one observation per line with a header. Our first step is to load the data and plot it

```
> X <- scan("KWhrs.dat", skip = 1)
> plot(X, type = "l")
```

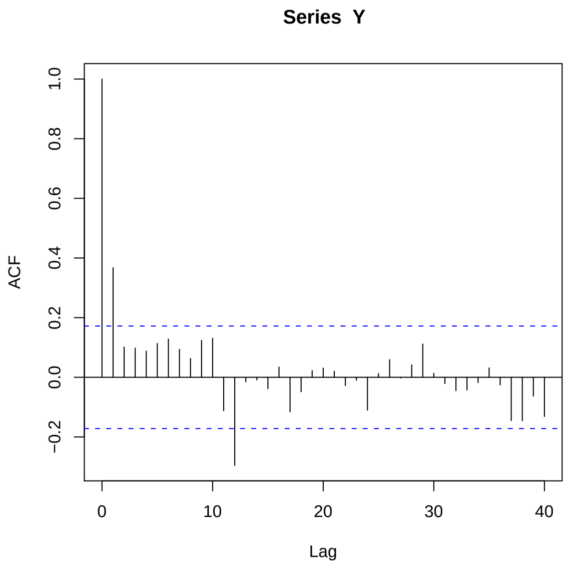


There is clearly a seasonal component. We remove this then calculate the acf and pacf.

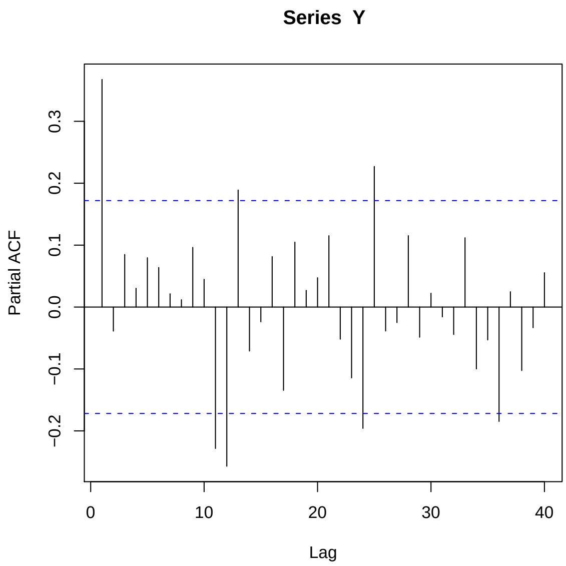
```
> n <- length(X)
> Y <- X[13:n] - X[1:(n - 12)]
> plot(Y, type = "l")
```



```
> acf(Y, lag = 40)
```




```
> pacf(Y, lag = 40)
```



The acf is consistent with stationarity. The acf has a cut-off after lag 12 but the pacf decays gradually, indicating an MA model would be appropriate. Spikes at lag 1 and 12 (but not in-between) indicate a seasonal MA component, whence we fit a SARIMA model of the form:

$$X_t \sim SARIMA\ (0,0,1) \times (0,1,1)_{12}$$

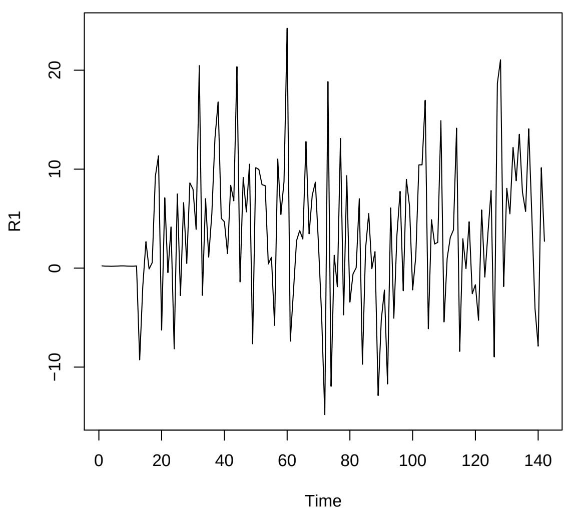
Note that we expect Y_t to have a mean, since X_t has a trend. We are supposing at the moment that the lag 12 differencing we used to remove the seasonal component was also sufficient to remove the trend.

```
> model1 <- arima(X, order = c(0, 0, 1), seasonal = list(order = c(0,
+     1, 1), period = 12), include.mean = T)
> model1
```

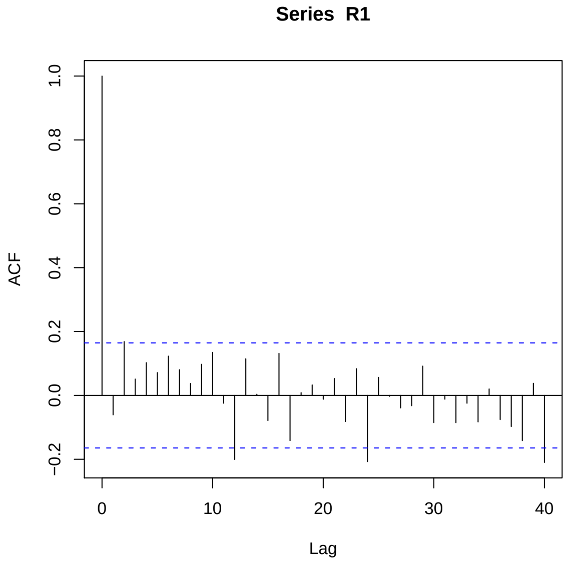
```
Call:
arima(x = X, order = c(0, 0, 1), seasonal = list(order = c(0, 1, 1), period = 12),
    include.mean = T)

Coefficients:
         ma1     sma1
      0.4746  -0.1583
s.e.  0.0678   0.0912

sigma^2 estimated as 70.62:  log likelihood = -461.47,  aic = 928.94
```

We check the fit of the model from the residuals

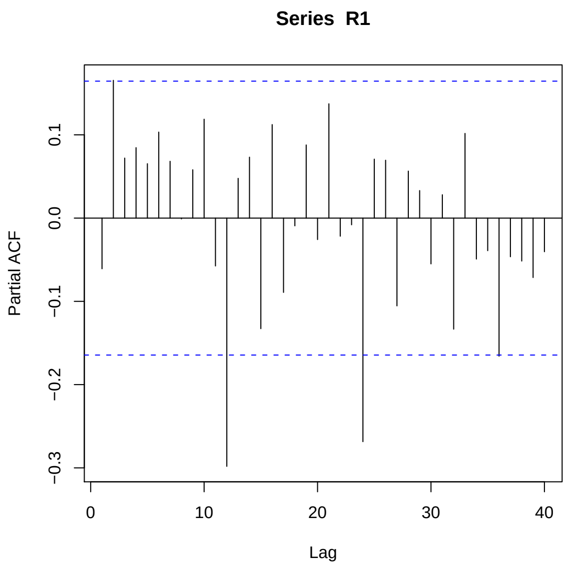
```
> R1 <- model1$residuals
> plot(R1, type = "l")
```




```
> acf(R1, lag = 40)
```

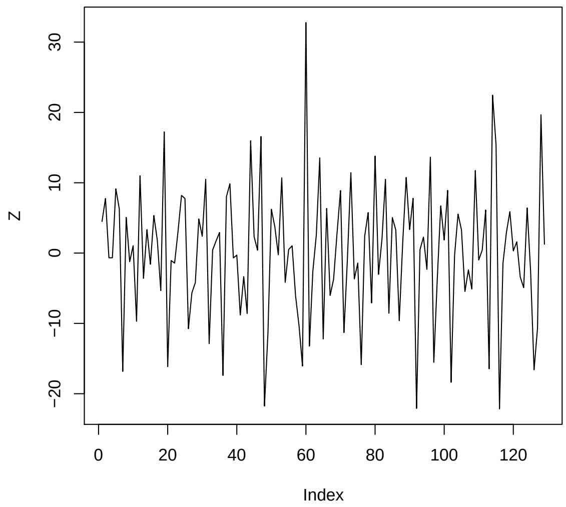



```
> pacf(R1, lag = 40)
```

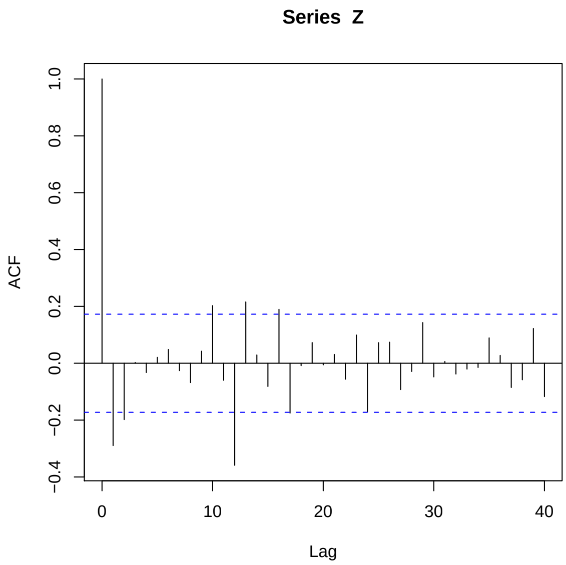


Both the acf and pacf show significant correlation remaining in the residuals, so we need a model of higher order. One could try increasing the order of the MA component or the seasonal MA component, adding an AR or seasonal AR component, or more differencing. Because the original data has a trend as well as seasonal component we try further differencing first:

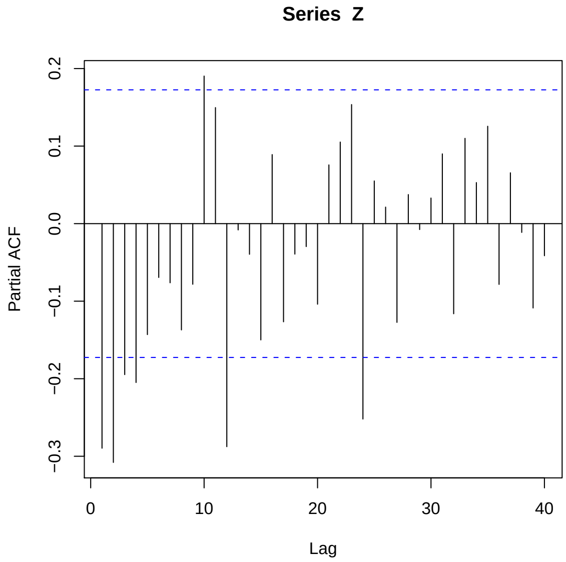
```
> nY <- length(Y)
> Z <- Y[2:nY] - Y[1:(nY - 1)]
> plot(Z, type = "l")
```



> *acf(Z, lag = 40)*



> *pacf(Z, lag = 40)*



The acf is consistent with stationarity. The acf has a cut-off after lag 16 but the pacf decays gradually, indicating an MA model would be appropriate. Spikes at lag 1, 2 then a bunch around 12 (but not in-between) indicate a seasonal MA component, whence we fit a SARIMA model of the form:

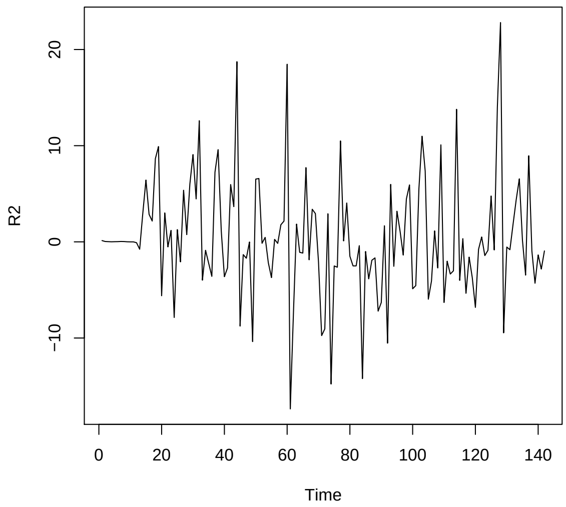
$$X_t \sim SARIMA\ (0,1,2) \times (0,1,1)_{12}$$

Note that we expect Z_t to have zero mean. This model can account for spikes at lags 1, 2, 10, 11, 12, 13 and 14; we are tacitly hoping that the spike at lag 16 is not important.

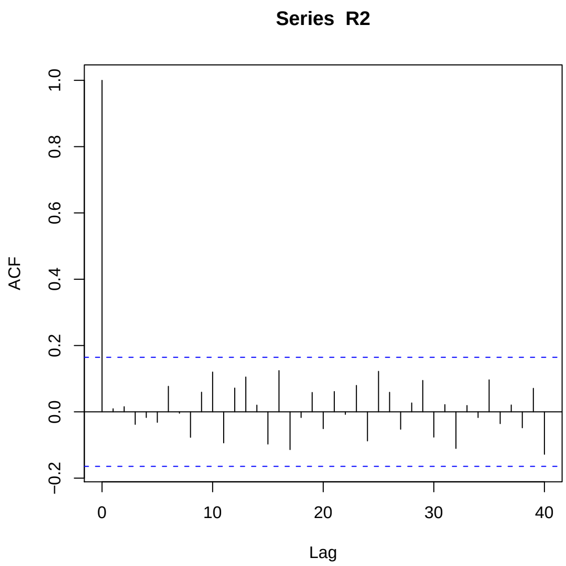
```
> model2 <- arima(X, order = c(0, 1, 2), seasonal = list(order = c(0,
+     1, 1), period = 12), include.mean = F)
> model2

Call:
arima(x = X, order = c(0, 1, 2), seasonal = list(order = c(0, 1, 1), period = 12),
    include.mean = F)

Coefficients:
          ma1      ma2     sma1
      -0.5755  -0.2298  -0.8053
s.e.   0.0852   0.0831   0.0938

sigma^2 estimated as 40.78:  log likelihood = -429.04,  aic = 866.08
```

We check the fit of the model from the residuals

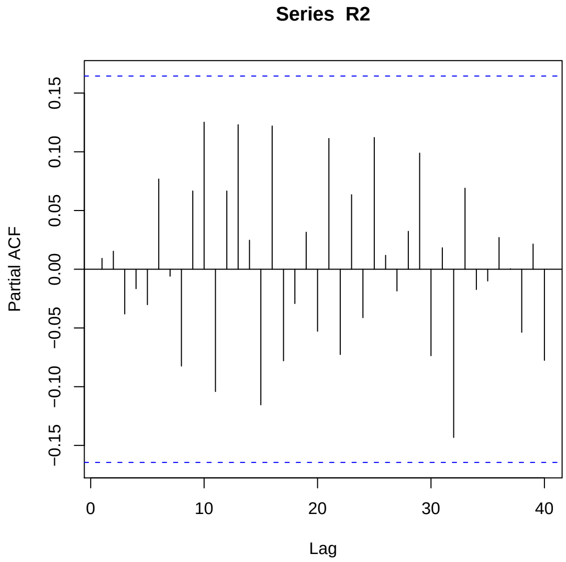
```
> R2 <- model2$residuals
> plot(R2, type = "l")
```



> acf(R2, lag = 40)



> pacf(R2, lag = 40)



This time the acf and pacf for the residuals are both consistent with white noise, so we accept the model and use it to forecast ahead 24 months.

```
> FC <- predict(model2, n.ahead = 24)
> plot(c(X, FC$pred), type = "l", main = "forecast", ylab = "kilowatt hours")
> lines(seq(n + 1, n + length(FC$pred)), FC$pred + 2 * FC$se, col = "red")
> lines(seq(n + 1, n + length(FC$pred)), FC$pred - 2 * FC$se, col = "red")
```