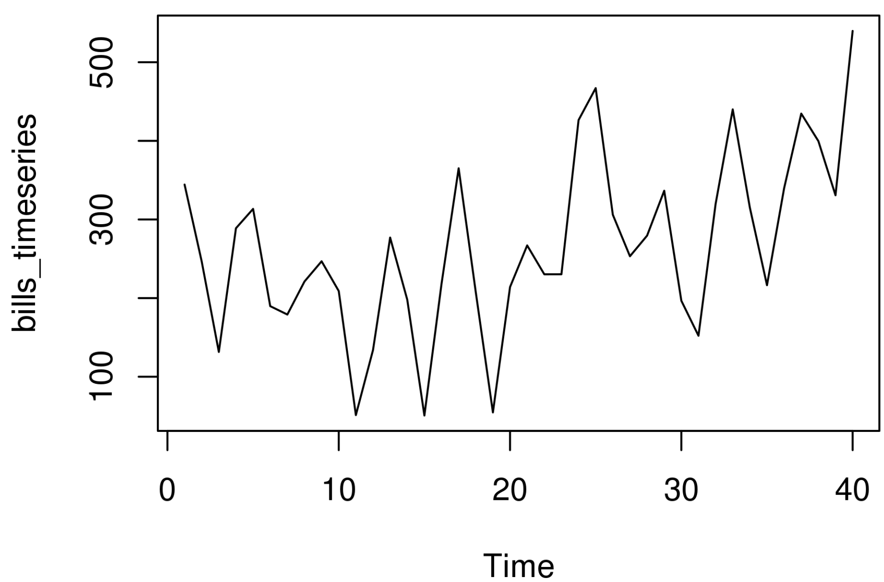
Homework Assignment 10

Matthew Tiger

November 19, 2015

Problem 1. Plot the energy bills versus time. What kind of trend appears to exist? What type of seasonal variation appears to exist? Is a transformation needed to obtain a series that displays constant variation?

Solution. See below for a plot of the bills time series data:



It is clear from the plot that there is a trend that appears to move upwards as time increases and a seasonal variation with period 4 lags present in the data so a transformation is needed to obtain residuals that represent a stationary time series. □

Problem 2. Write algebraically a time series model with trend and seasonal component with definitions of the dummy variable.

Solution. Note that it appears that this time series has a linear trend. Additionally, we are interested in capturing the seasonal quarter data of the time series. Therefore, a time series model for the data with trend and seasonal components is given by

$$X_t = a_0 + a_1 t + a_2 Q_1 + a_3 Q_2 + a_4 Q_3 + a_5 Q_4 + Y_t,$$

where we define Q_i as 1 if $t \equiv i \mod 4$ and 0 otherwise and a_j is constant and Y_t is a time series model. □

Problem 3. Are all the variables in the model statistically significant? Justify your answer.

Solution. The following R code performs a linear regression on our data set using the above equation:

```
quarter_variable <- function(ts, position){
    vector <- rep(0, 4)
    vector[position] <- 1

    variable <- rep(vector, length(ts) / 4)

    return(variable)
}

bills <- scan("bills.csv", skip=1)
bills.ts <- ts(bills)

bills.ts.Q1 <- quarter_variable(bills.ts, 1)
bills.ts.Q2 <- quarter_variable(bills.ts, 2)
bills.ts.Q3 <- quarter_variable(bills.ts, 3)
bills.ts.Q4 <- quarter_variable(bills.ts, 4)

bills.ts.regression_equation <- bills.ts ~ 0 + time(bills.ts) +
    bills.ts.Q1 + bills.ts.Q2 + bills.ts.Q3 + bills.ts.Q4
bills.ts.regression <- lm(bills.ts.regression_equation)

# The following tells us that all variables are significant
# using a significance level of alpha = 0.05.
summary(bills.ts.regression)
```

The code above outputs the following table displaying the significance of the variables in the regression equation:

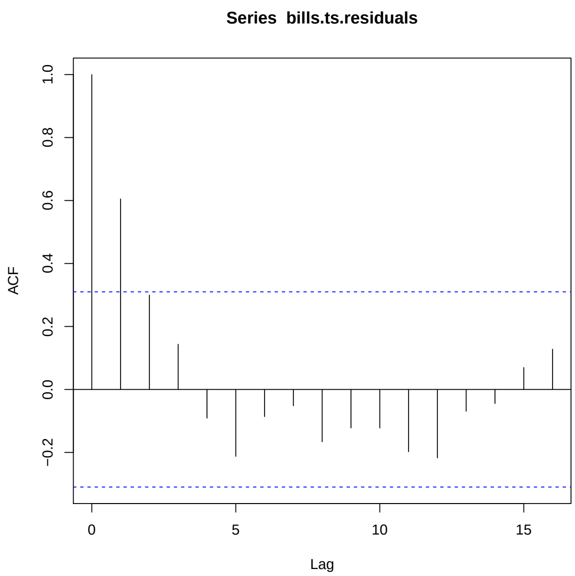
```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
time(bills.ts)    4.8922     0.9753   5.016 1.53e-05 ***
bills.ts.Q1     256.2549    29.0803   8.812 2.08e-10 ***
```

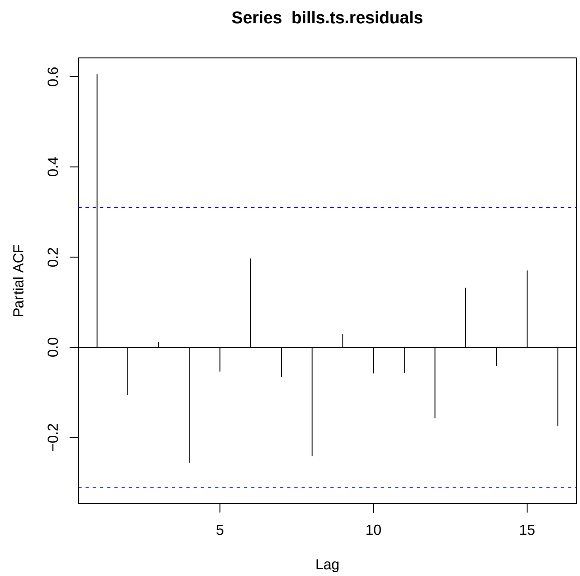
```
bills.ts.Q2    152.0117    29.7113   5.116 1.13e-05 ***
bills.ts.Q3     62.2705    30.3606   2.051   0.0478 *
bills.ts.Q4    190.4843    31.0269   6.139 5.06e-07 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

From this table we see that all of our of our variables are statistically significant using a significance level of $\alpha = 0.05$. $\square$

Problem 4. Plot the ACF and PACF of the residual. Are the residuals correlated?

Solution. The following are the ACF and PACF plots of the residuals of our time series and our regression equation.





Note that the residuals are correlated as the ACF shows a nonzero value for a lag greater than 0. $\square$

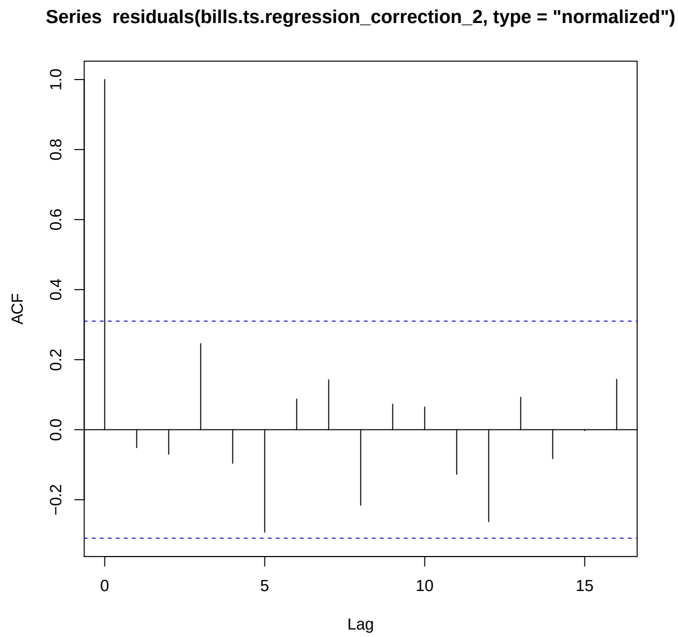
Problem 5. What is the time series model that best fits the model? Is the model statistically significant?

Solution. Note that are residuals are not white noise so we need to rerun the regression analysis using the fact that the residuals are correlated as an ARMA(1,1) model.

Rerunning the regression with this fact shows that the Q_3 variable is no longer significant. Removing this variable and rerunning the analysis shows that

```
                    Value Std.Error   t-value p-value
time(bills.ts)    7.44860  1.157879  6.432974       0
bills.ts.Q1     206.68710 20.346380 10.158421       0
bills.ts.Q2      95.32084 16.141974  5.905154       0
bills.ts.Q4     130.94859 16.143011  8.111782       0
```

verifying that all of our variables are now statistically significant. The ACF of the residuals is shown below:



From the graph, it is clear that with our new regression equation the residuals are white noise. The underlying noise of the regression is given by the ARMA(1,1) process $\{Y_t\}$ with $Y_t = 0.6055Y_{t-1} + Z_t + 0.3097Z_{t-1}$ with $Z_t \sim \mathrm{WN}(0, 75.3106^2)$.

Therefore, a statistically significant time series model is given by

$$X_t = 7.45t + 206.68Q_1 + 95.32Q_2 + 130.95Q_4 + 0.61Y_{t-1} + Z_t + 0.31Z_{t-1},$$

where $Z_t \sim \mathrm{WN}(0, 75.3106^2)$. $\square$