STAT 223: Project 4

Nrepesh Joshi

Professor Kevin Hastings

Applied Analytics

5/14/20

Home energy consumption

The purpose here is to detect any trends or other patterns in home energy consumption, and to enable me to identify baseline gas and electrical usage, as distinct from seasonally dependent usage. What usage statistics are predicted to occur in 2019?

1. Title: Home Energy Usage (energy.csv)
2. Relevant Information:

These data comprise 4 years of home energy usage in a suburban Galesburg ranch home built in 1969 from January of
2015 through December of 2018. The home has the usual electrical appliances, gas heat and hot water, and electrically
driven central air conditioning. A happily married husband and wife live in the home, the former of which is always too cold
while the latter is always too hot. From January of 2015 through July of 2016, a college student (my daughter Emily) lived
in the home. She takes long, hot showers.

3. Number of Instances: 48.
4. Number of Attributes: 3, including the month
5. Attribute Information:

1. Month: character
2. Elec (kwh): numeric
3. Gas (therms): numeric

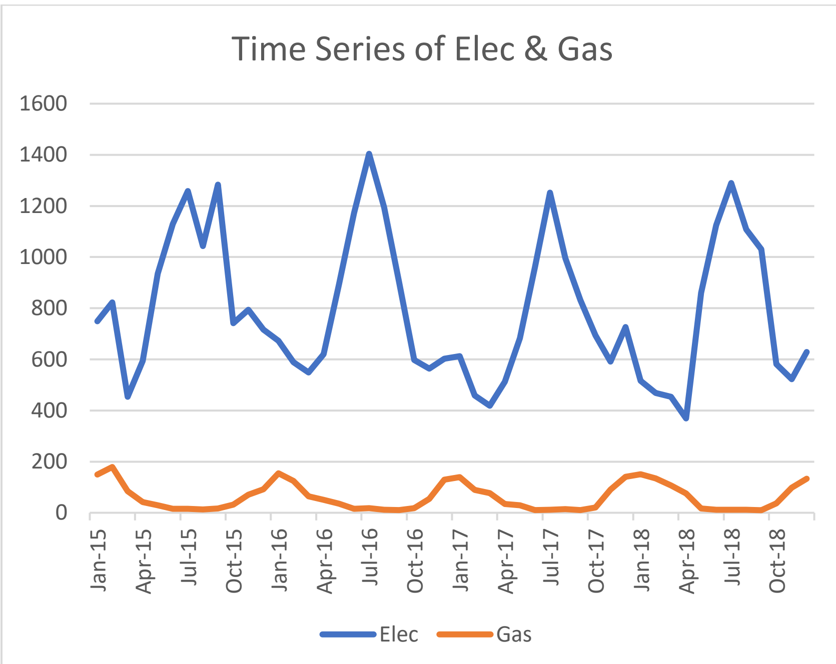
6. Missing Attribute Values: none.

==============================

DATA ANALYSIS

==============================

1. The data file is in Classroom for download under the name "energy.csv". First, load it into Excel and save as an Excel file. Produce a time series plot of the electric and gas series on the same graph. Describe any patterns that you see and speculate on reasons for those patterns. Copy the Excel graph into your project write-up.



For both the electrical and gas time series, we can see a seasonal aspect where the rate increases periodically. We see spikes in the electric bill during the summer months (Apr-Jul). This might be because the air conditioning is required throughout the day because the temperature is too warm outside. We can also see peaks in the gas bill during the winter months (Oct-Jan). This might be because heating is required more in colder seasons.

We do not see an increasing or decreasing trend as time flows. We cannot certainly say about the random aspect either as it requires more analysis later in this exercise.

2. For the electric data only, create columns in your spreadsheet that contain the naive forecasts, average forecasts, and moving average order 3 forecasts. For each of these forecasts, create a column for the absolute errors, and compute at the bottom the mean absolute error. Copy the table into your project write-up as an appendix, and report on which forecasting method was most successful according to the MAE criterion.

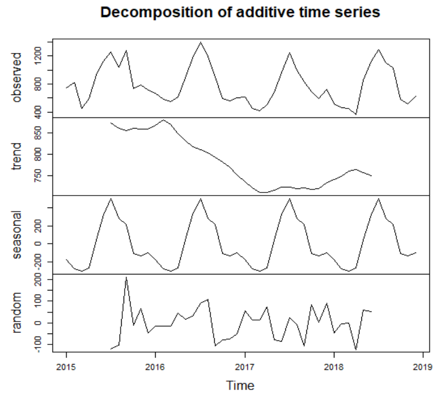
= Table attached in Appendix. From our excel calculation, we see that the MAE for naïve forecast is 175.83, MAE for average forecast is 257.54 and MAE for moving average forecast is 243.207. From the MAE criterion, we are trying to reduce the errors in our time series forecast so we can conclude here that naïve is the best forecasting method for our time series. This means that the bill for electricity depends on the previous month.

3. Now load the "energy.csv" file into R and create time series objects for both electric and gas. Decompose each series into a trend, seasonal, and random component with the decompose (seriesname) command as shown in the notes file. For each, do a regression of the trend data on time to check whether there is a significant linear increase or decrease in energy usage. Given your finding about the trend, and the pattern of the seasonal component, what would you predict for my electrical and gas usages for the first three months of 2019?

= *For electricity*

Below is the decomposition and plot for electricity time series.

```
> decompElec <- decompose(electimeseries)
> decompElec$trend
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov      Dec
2015       NA       NA       NA       NA       NA       NA 873.4583 860.5000 854.7083 859.7500 859.0417 859.0417
2016 866.9167 879.3333 869.7500 847.8333 832.2500 817.9167 810.6667 802.7083 791.7917 781.8333 768.5417 750.9167
2017 735.7500 721.0833 709.8333 710.9583 716.1250 722.4583 723.6667 720.0833 721.9583 717.4583 718.9167 733.1667
2018 741.5833 747.8750 760.9167 764.5417 756.9583 750.0000       NA       NA       NA       NA       NA       NA
> decompElec$seasonal
            Jan        Feb        Mar        Apr        May        Jun        Jul        Aug        Sep        Oct        Nov        Dec
2015 -180.15856 -276.50579 -305.57523 -272.85301   44.48032  323.13310  503.32755  284.49421  216.77199 -107.75579 -131.24190  -98.11690
2016 -180.15856 -276.50579 -305.57523 -272.85301   44.48032  323.13310  503.32755  284.49421  216.77199 -107.75579 -131.24190  -98.11690
2017 -180.15856 -276.50579 -305.57523 -272.85301   44.48032  323.13310  503.32755  284.49421  216.77199 -107.75579 -131.24190  -98.11690
2018 -180.15856 -276.50579 -305.57523 -272.85301   44.48032  323.13310  503.32755  284.49421  216.77199 -107.75579 -131.24190  -98.11690
> decompElec$random
             Jan         Feb         Mar         Apr         May         Jun         Jul         Aug         Sep         Oct         Nov
2015          NA          NA          NA          NA          NA          NA -118.785880 -101.994213  211.519676  -10.994213   66.200231
2016  -14.758102  -13.827546  -15.174769   45.019676   16.269676   30.950231   90.005787  107.797454 -107.563657  -77.077546  -73.299769
2017   56.408565   13.422454   13.741898   73.894676  -78.605324  -85.591435   25.005787   -9.577546 -107.730324   84.297454    3.325231
2018  -45.424769   -3.369213   -2.341435 -122.688657   58.561343   50.866898          NA          NA          NA          NA          NA
             Dec
2015  -44.924769
2016  -50.799769
2017   91.950231
2018          NA
```



Regression and plot of the trend data for electricity

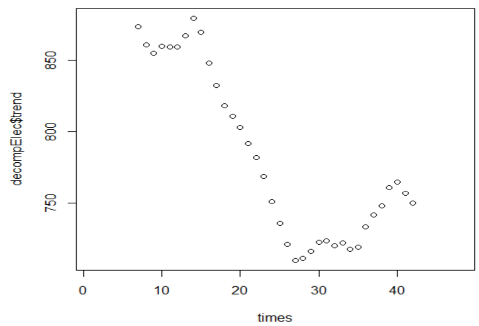
```
> elecmodel <- lm(decompElec$trend~times); summary(elecmodel)

Call:
lm(formula = decompElec$trend ~ times)

Residuals:
   Min     1Q Median     3Q    Max 
-60.63 -22.27   1.60  20.02  55.24 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 897.4979    14.1449  63.450  < 2e-16 ***
times        -4.7048     0.5315  -8.851 2.41e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 33.13 on 34 degrees of freedom
  (12 observations deleted due to missingness)
Multiple R-squared:  0.6974,	Adjusted R-squared:  0.6885 
F-statistic: 78.35 on 1 and 34 DF,  p-value: 2.409e-10
```



We see here that the slope is -4.7048 so that means that there is a significant linear decrease in the electrical uses. The R^2 is about 0.7 which tells us that it is a pretty good model of the electric trend.

For prediction of next 3 electricity bills (49,50,51)

```
> elecpredict <- lm(dset$Elec~times+month); summary(elecpredict)

Call:
lm(formula = dset$Elec ~ times + month)

Residuals:
     Min       1Q   Median       3Q      Max 
-156.613  -69.181   -0.975   61.656  199.838 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  712.894     51.604  13.815 9.79e-16 ***
times         -3.981      1.028  -3.874 0.000449 ***
month2       -48.769     67.560  -0.722 0.475178    
month3      -160.787     67.584  -2.379 0.022941 *  
month4      -101.556     67.623  -1.502 0.142116    
month5       221.425     67.677   3.272 0.002407 ** 
month6       478.906     67.748   7.069 3.11e-08 ***
month7       687.638     67.833  10.137 5.93e-12 ***
month8       475.869     67.934   7.005 3.77e-08 ***
month9       406.100     68.051   5.968 8.52e-07 ***
month10       51.831     68.183   0.760 0.452237    
month11       20.313     68.330   0.297 0.768016    
month12       75.044     68.492   1.096 0.280710    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 95.53 on 35 degrees of freedom
Multiple R-squared:  0.9123,	Adjusted R-squared:  0.8823 
F-statistic: 30.35 on 12 and 35 DF,  p-value: 6.108e-15
```

The model has an R^2 value of 0.9 which means that is a very good predictor of our electricity bill.

For January (electric):

= 712.894 -3.981*49-(0)   [uses dummy variable so for first month everything else is 0]

= 517.85

For February (electric):

= 712.894 -3.981*50 - 48.769*1-0-0…….

= 465.075

For March (electric):

= 712.894 -3.981*51 – 160.787*1-0-0…….

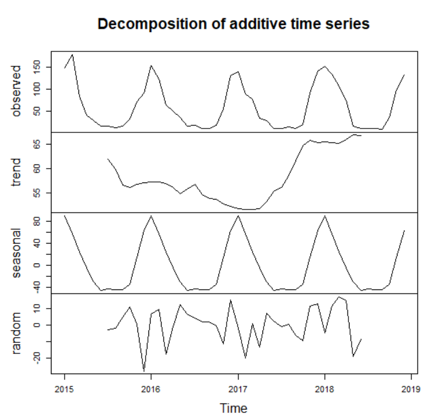
= 349.076

Hence the prediction for the next 3 electric bills are $517.85, $465.075, $349.076

For gas

Below is the decomposition and plot for gas time series.

```
> decompGas <- decompose(gastimeseries)
> decompGas$trend
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov      Dec
2015       NA       NA       NA       NA       NA       NA 61.95833 59.87500 56.70833 56.20833 56.87500 57.16667
2016 57.25000 57.29167 57.00000 56.16667 54.91667 55.83333 56.83333 54.79167 53.87500 53.70833 52.70833 52.20833
2017 51.75000 51.58333 51.66667 51.75000 53.33333 55.29167 56.20833 58.54167 61.66667 64.66667 65.91667 65.45833
2018 65.50000 65.41667 65.29167 65.95833 66.95833 66.91667       NA       NA       NA       NA       NA       NA
> decompGas$seasonal
           Jan       Feb       Mar       Apr       May       Jun       Jul       Aug       Sep       Oct       Nov       Dec
2015  90.14583  57.54861  24.65972  -4.31250 -31.09028 -46.36806 -43.02083 -44.75694 -44.43750 -34.88194  13.81250  62.70139
2016  90.14583  57.54861  24.65972  -4.31250 -31.09028 -46.36806 -43.02083 -44.75694 -44.43750 -34.88194  13.81250  62.70139
2017  90.14583  57.54861  24.65972  -4.31250 -31.09028 -46.36806 -43.02083 -44.75694 -44.43750 -34.88194  13.81250  62.70139
2018  90.14583  57.54861  24.65972  -4.31250 -31.09028 -46.36806 -43.02083 -44.75694 -44.43750 -34.88194  13.81250  62.70139
> decompGas$random
            Jan         Feb         Mar         Apr         May         Jun         Jul         Aug         Sep         Oct         Nov         Dec
2015         NA          NA          NA          NA          NA          NA  -2.9375000  -2.1180556   4.7291667  10.6736111   0.3125000 -27.8680556
2016  6.6041667   9.1597222 -17.6597222  -0.8541667  12.1736111   6.5347222   4.1875000   1.9652778   1.5625000  -0.8263889 -11.5208333  15.0902778
2017 -1.8958333 -20.1319444   0.6736111 -13.4375000   6.7569444   2.0763889  -1.1875000   0.2152778  -6.2291667  -9.7847222  11.2708333  12.8402778
2018 -4.6458333  11.0347222  17.0486111  14.3541667 -18.8680556  -8.5486111          NA          NA          NA          NA          NA          NA
```



Regression and plot of the trend data for gas

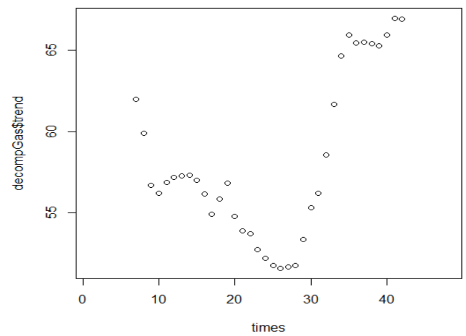
```
> gasmodel <- lm(decompGas$trend~times); summary(gasmodel)

Call:
lm(formula = decompGas$trend ~ times)

Residuals:
   Min     1Q Median     3Q    Max 
-7.355 -3.715  1.261  3.737  8.274 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 51.87695    1.86630  27.797   <2e-16 ***
times        0.25814    0.07013   3.681    8e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.371 on 34 degrees of freedom
  (12 observations deleted due to missingness)
Multiple R-squared:  0.2849,    Adjusted R-squared:  0.2639 
F-statistic: 13.55 on 1 and 34 DF,  p-value: 0.0007999
```



We see here that the slope is 0.25814 so that means that there is a significant linear increase in the gas uses. However, the R^2 is only about 0.3 which tells us that it is not a good model of the gas trend.

For prediction of next 3 electricity bills (49,50,51)

```
> gaspredict <- lm(dset$Gas~times+month); summary(gaspredict)

Call:
lm(formula = dset$Gas ~ times + month)

Residuals:
    Min      1Q  Median      3Q     Max 
-43.142  -6.075   0.592   6.267  49.425 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 146.4681     8.7597  16.721  < 2e-16 ***
times         0.1069     0.1745   0.613   0.5438    
month2      -17.1069    11.4682  -1.492   0.1447    
month3      -65.4639    11.4721  -5.706 1.88e-06 ***
month4      -98.0708    11.4788  -8.544 4.40e-10 ***
month5     -121.1778    11.4881 -10.548 2.06e-12 ***
month6     -135.2847    11.5000 -11.764 1.03e-13 ***
month7     -134.6417    11.5145 -11.693 1.22e-13 ***
month8     -136.4986    11.5317 -11.837 8.62e-14 ***
month9     -137.1056    11.5515 -11.869 7.98e-14 ***
month10    -122.7125    11.5738 -10.603 1.79e-12 ***
month11     -70.8194    11.5988  -6.106 5.60e-07 ***
month12     -25.6764    11.6263  -2.208   0.0339 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16.22 on 35 degrees of freedom
Multiple R-squared:  0.9263,    Adjusted R-squared:  0.901 
F-statistic: 36.64 on 12 and 35 DF,  p-value: 3.17e-16
```

The model has an R^2 value of 0.9 which means that is a very good predictor of our gas bill.

For January (gas):

= 146.4681-0.1069*49-(0)   [uses dummy variable so for first month everything else is 0]

= 146.3612

For February (gas):

= 146.4681-0.1069*50-17.1069-0-0…..

= 124.0162

For March (gas):

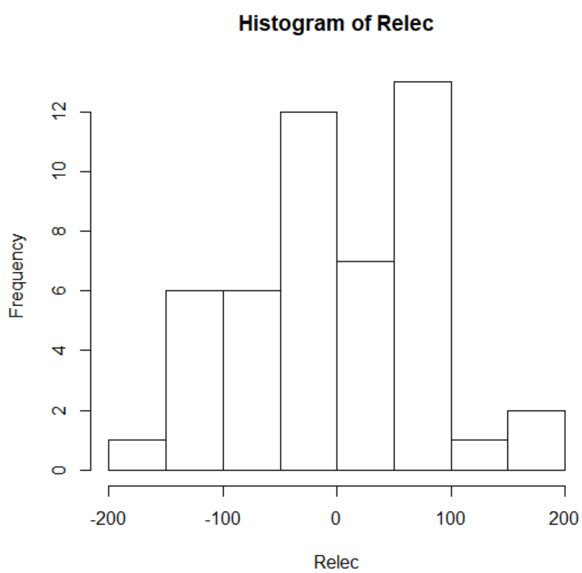
= 146.4681-0.1069*51-65.4639-0-0…..

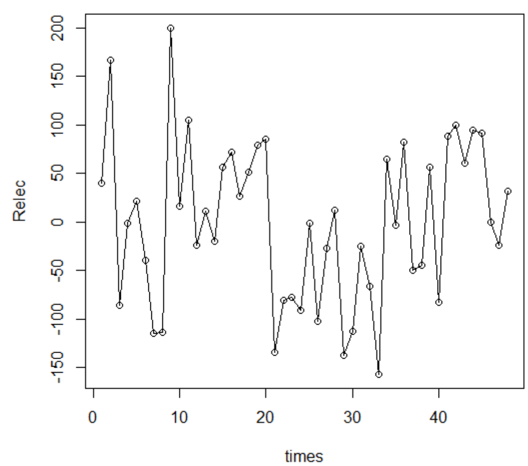
= 75.5523

Hence the prediction for the next 3 electric bills are $146.3612, $124.0162, $75.5523

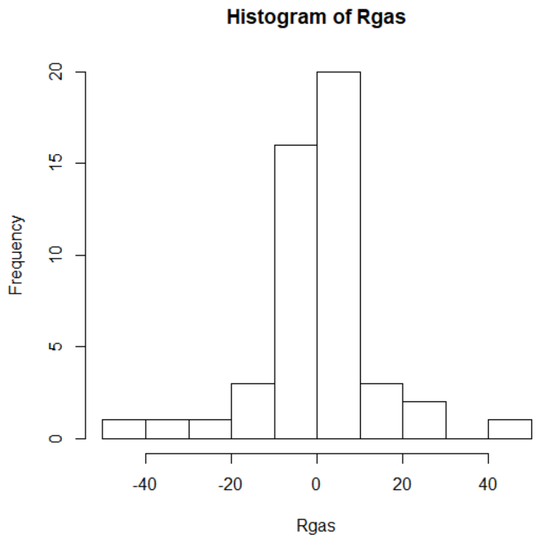
4. Lastly, look at the random components of each series. Do they appear to be stationary, and have mean 0? For each, form a series of the first 47 observations and the last 47 (i.e. the original series and the lagged series) and check for statistically significant correlation between Xt and Xt+1 to see whether the random component is just white noise.

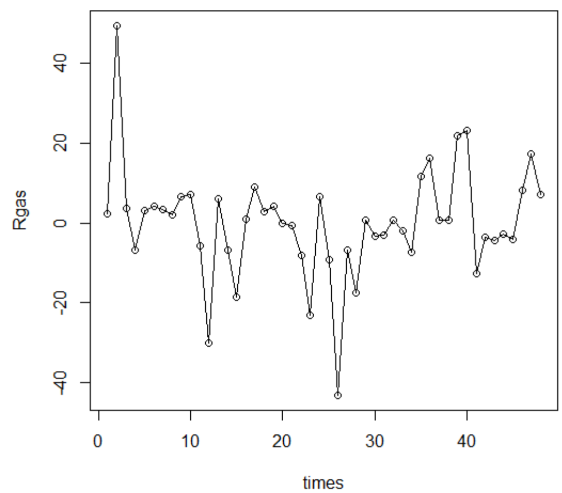
For electricity





For gas





The residuals for both electricity and gas appear to be normally distributed with mean of 0 and constant variance. The time plot does not show any patters either.

Random Noise test

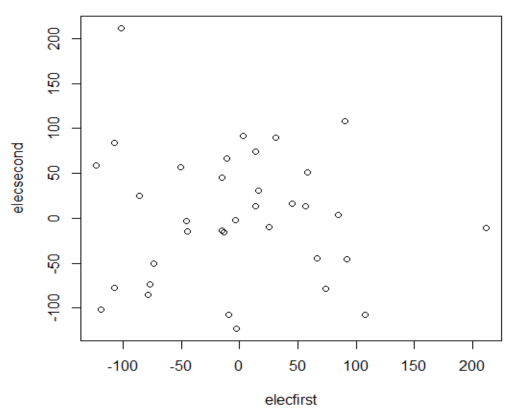
For electricity

```
Call:
lm(formula = elecfirst ~ elecsecond)

Residuals:
     Min       1Q   Median       3Q      Max 
-123.310  -59.253   -3.769   54.352  213.352 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.60070   12.98760   -0.20    0.843
elecsecond  -0.06986    0.17898   -0.39    0.699

Residual standard error: 76.8 on 33 degrees of freedom
  (12 observations deleted due to missingness)
Multiple R-squared:  0.004595,  Adjusted R-squared:  -0.02557 
F-statistic: 0.1523 on 1 and 33 DF,  p-value: 0.6988
```



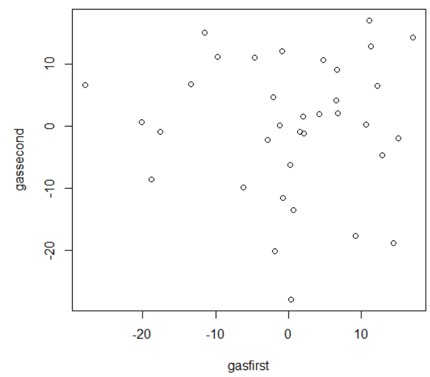
For Gas

```
Call:
lm(formula = gasfirst ~ gassecond)

Residuals:
     Min       1Q   Median       3Q      Max 
-28.0159  -3.9784   0.1623   7.5494  17.0413 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.26759    1.86155   0.144    0.887
gassecond   -0.01814    0.17262  -0.105    0.917

Residual standard error: 11.01 on 33 degrees of freedom
  (12 observations deleted due to missingness)
Multiple R-squared:  0.0003344,	Adjusted R-squared:  -0.02996 
F-statistic: 0.01104 on 1 and 33 DF,  p-value: 0.917
```



For both electricity and gas, when we form a series of the first 47 observations and the last 47 (i.e. the original series and the lagged series) and check for statistically significant correlation between Xt and Xt+1 we do not see any statistically significant correlation as the R^2 values are only slightly greater than 0. Electricity: 0.0045 and Gas: 0.00033. We can conclude that the random component is just white noise.

Appendix:

Graph file attached with assignment.

Code:

```
#-------------------------------------------------------
# Part 1 Importing data & packages

install.packages("TTR")
library("TTR")
dset <- read.csv("energy.csv", header = TRUE); View(dset)
head(dset)

#------------------------------------------------------------------------------------------

# Part 3 Creating Time Serries
```

```
# For electricity
electimeseries <- ts(dset$Elec, frequency = 12, start=c(2015,1))
plot.ts(electimeseries)

decompElec <- decompose(electimeseries)
decompElec$trend
decompElec$seasonal
decompElec$random
plot(decompElec)

#For Gas
gastimeseries <- ts(dset$Gas, frequency = 12, start=c(2015,1))
plot.ts(gastimeseries)

decompGas <- decompose(gastimeseries)
decompGas$trend
decompGas$seasonal
decompGas$random
plot(decompGas)

#-------------------------------------------------------------------------------------------

# Linear Regression

month<- factor(rep(c(1,2,3,4,5,6,7,8,9,10,11,12),4))
times <- c(1:48)

# For electricity
elecmodel <- lm(decompElec$trend~times); summary(elecmodel)
plot(times, decompElec$trend)


# For gas
```

```
gasmodel <- lm(decompGas$trend~times); summary(gasmodel)
plot(times, decompGas$trend)


# For prediction
elecpredict <- lm(dset$Elec~times+month); summary(elecpredict)

gaspredict <- lm(dset$Gas~times+month); summary(gaspredict)


#----------------------------------------------------------------------------------------

# Part 4 Random components

Relec <- elecpredict$residuals
hist(Relec)
plot(times,Relec,type = 'n')
lines(times,Relec,type='o')

Rgas <- gaspredict$residuals
hist(Rgas)
plot(times,Rgas,type = 'n')
lines(times,Rgas,type='o')

#----------------------------------------------------------------------------------------
# White Noise Test

elecresu <- decompElec$random[1:48]
elecfirst<- elecresu[1:47]
elecsecond<- elecresu[2:48]
elecresidsmodel<-lm(elecfirst~elecsecond);summary(elecresidsmodel)
plot(elecfirst,elecsecond)
```

```
gasresu <- decompGas$random[1:48]
gasfirst<- gasresu[1:47]
gassecond<- gasresu[2:48]
gasresidsmodel<-lm(gasfirst~gassecond); summary(gasresidsmodel)
plot(gasfirst,gassecond)

#-----------------------------------------------------------------------------------------
```