Statistics 626 - Assignment 03

Vivek Gupta*

Department of Statistics, Texas A & M University

June 9, 2018

*vivek235@tamu.edu

1 I

1.1 a.

The two random walks x_t , y_t are simulated and shown below in the Figure 1 These are simulated with initial values of x_0 , $y_0 = 0$, using two independent N(0,1) white noises.

1.2 i.

The plots are shown in Figure 2 and 3. The first plot shows the lag 0 relationship between series x and y. While the second plot shows the relationship between x_{t-h} and y_t at each of the lags from $h = 0, 1, 2, 3 \ldots 10$

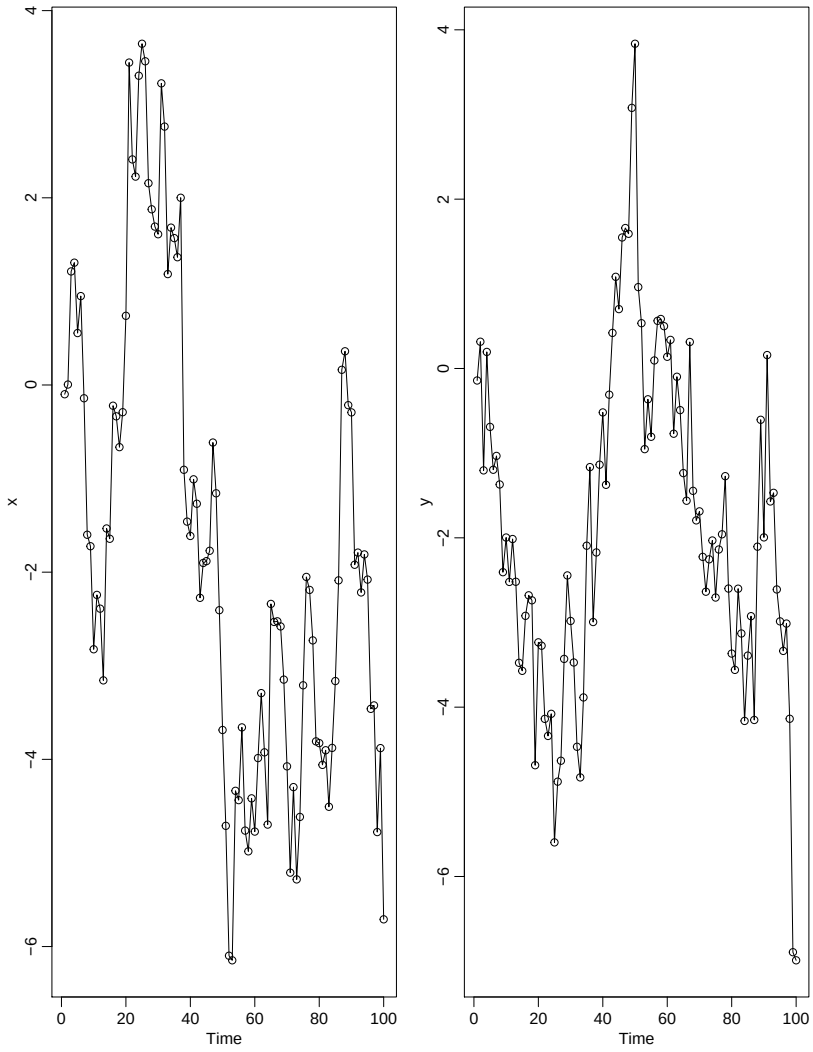

Figure 1: Input series , y_t vs x_t

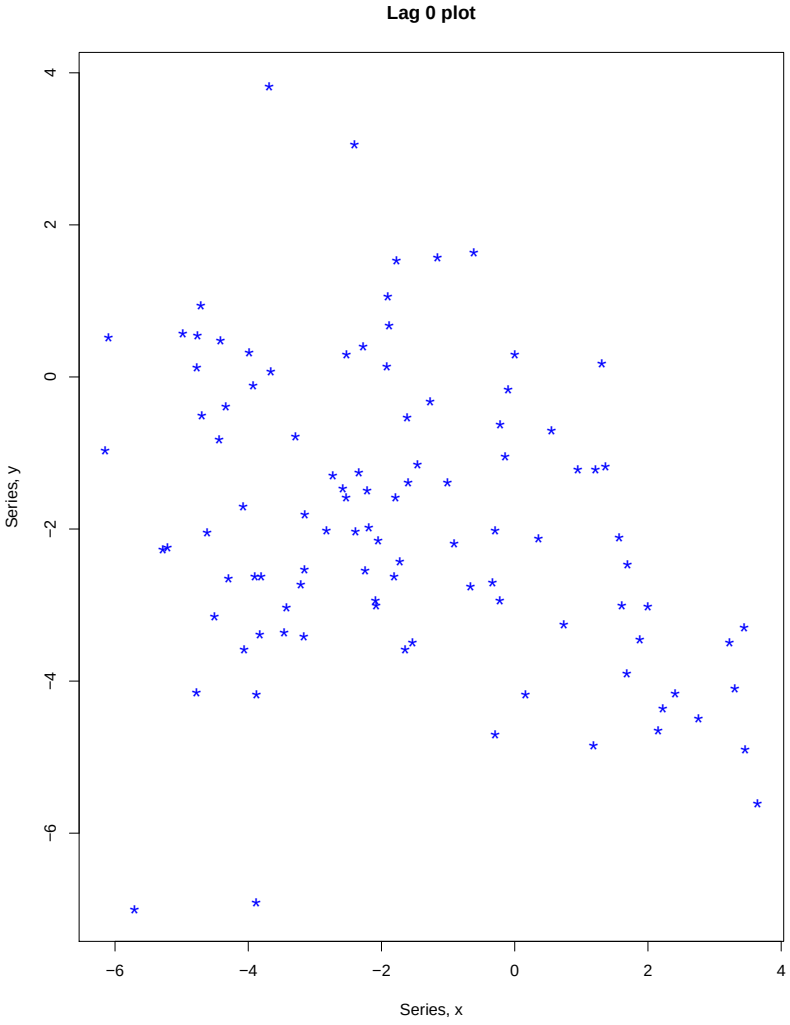

Figure 2: Lag 0 plot , y_t vs x_t

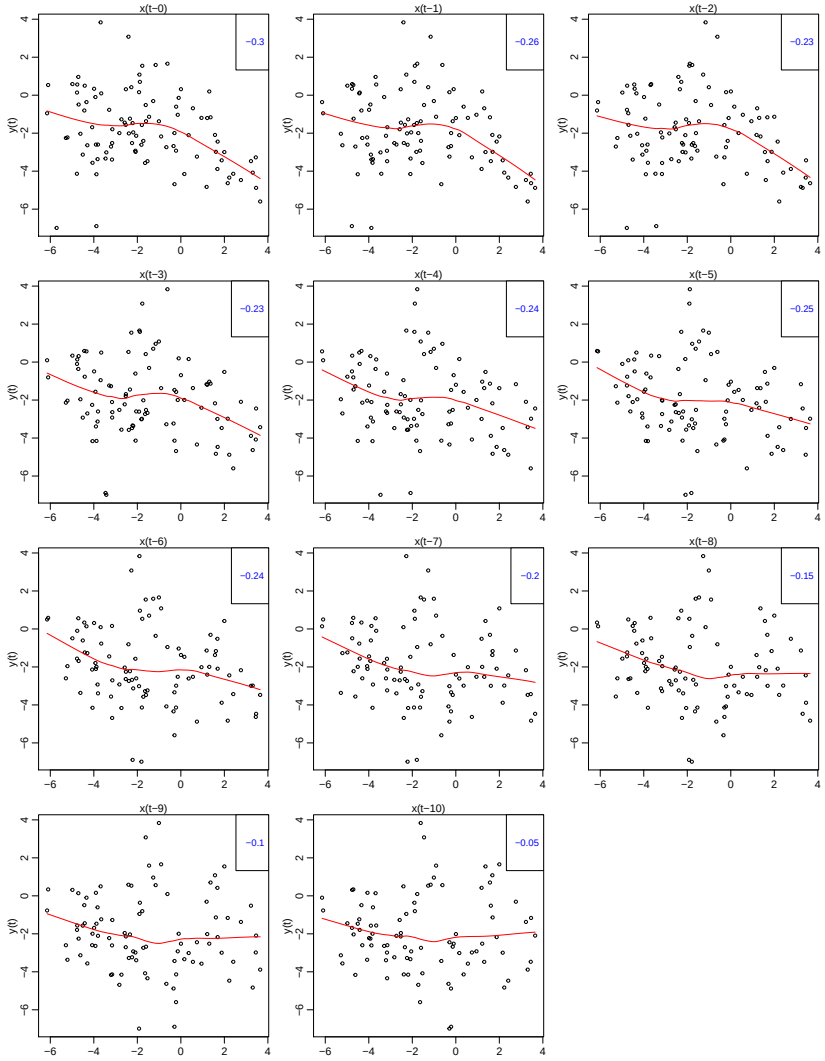

Figure 3: Lag h plot , y_t vs x_{t-h}

We notice some non - linear relationship between the two series as shown in the above figure as is evident from the smoothed fit in each of the graphs. This relationship differs subtly at various lags. At lag valeu $h = 9$, 10 we fail to notice any relationship between the two series.

The correlation coefficient as shown is all less than 0.5 and hence we do not have a very strong linear relationship between these two series at various values of the lag.

1.3 ii.

Consider the linear regression between x_t and y_t as shown below

$$y_t = \beta_0 + \beta_1 x_t + w_t$$

The above, simple linear regression model is valid upon certain constraints or assumptions that w_t is $i.i.d$ N(0,σ^2)

But since there isn't a linear relationship as evident from Figure 3 , we expect to fail to reject the null hypothesis.

1.4 iii.

The linear fit summary is shown in Figure 4 . The test for

$$H_o : \beta_1 = 0$$

is significant given the output of the model at a p-value of 0.00237 and

$$\alpha = 0.05$$

.

When we look at the residual plot as in Figure 5 to diagnose the validity of the model we notice that the residuals does not follow the assumption of the model. The residuals vs. the fitted definitely shows a non linear pattern and the residuals do not look to be coming from a normal distribution. Thus, the assumptions of the model are not met with this inear fit.

However, if we assume that the model is valid , then the we have evidence to reject the null hypothesis and conclude that at p-value < 0.05, the covariate x_t is significant in explaining the variability in y_t.

```
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-6.0868 -1.2320 -0.2089  1.3436  5.2187

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.25842    0.22731  -9.935  < 2e-16 ***
x           -0.23738    0.07606  -3.121  0.00237 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.881 on 98 degrees of freedom
Multiple R-squared:  0.0904,    Adjusted R-squared:  0.08112
F-statistic: 9.739 on 1 and 98 DF,  p-value: 0.00237
```

Figure 4: Linear fit between x_t , y_t

1.5 b

We repeated the above experiment 1000 times and noted that approx. 75 pc. of the times the test concluded that the covariate x_t is significant.

Clearly this does not support our expectation based on what we noted for the relationship in Figures 2 and 3 .

We also has noted earlier based on residual plot shown in Figure. 5 that the model is invalid due to

1. The relationship between x_t and y_t is non linear.
2. Residuals show a pattern of non linearity and thus are not independent.
3. We have evidence of heteroscedasticity in the model.
4. Normality in the residuals is questionable.

Since the model is invalid , the statistical summaries and results such as the p-value of the AOV and significance of the covariate are all invalid and thus the result of linear fit cannot be relied upon.

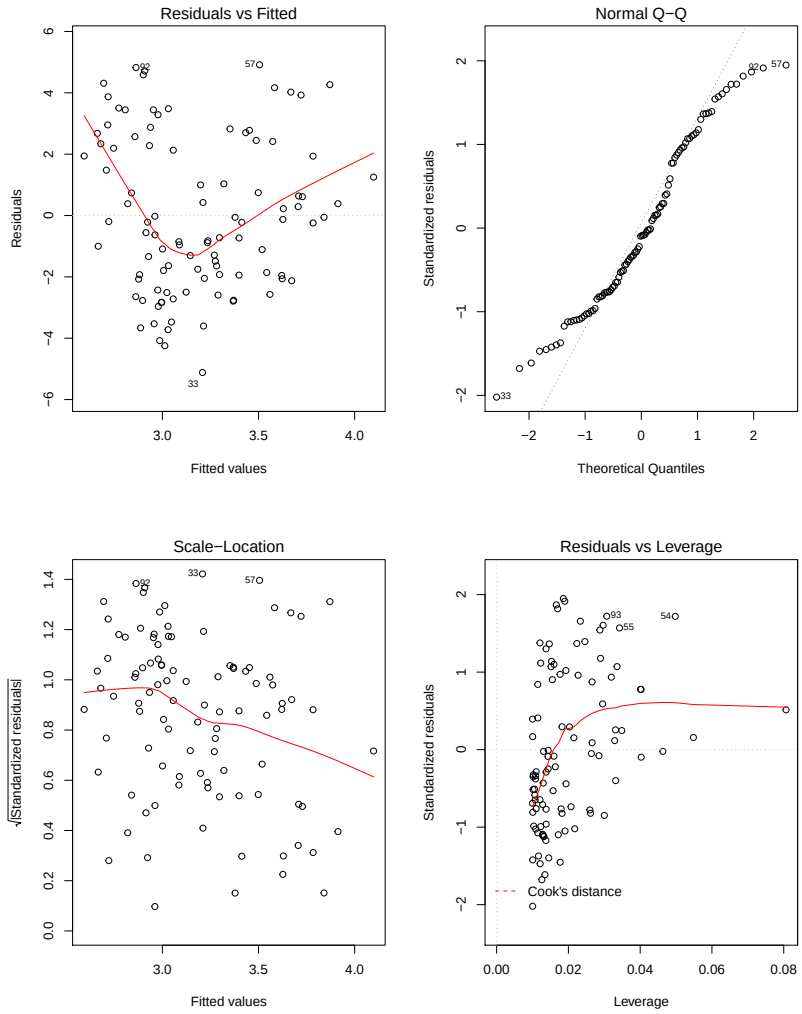

Figure 5: Residual Plot of the linear fit

2 2.6

2.1 a

We can notice the evidence of non homogeneous variances in the varve observations in Figure 6 . We divide the dataset into two equal halves and plot them as shown in the Figure 9.Difference in the variance is clearly visible in the two halves of the dataset. We can see the range of the varve varies significantly in the two halves.

We then take a log transformation of the data to ascertain is the variances are controlled by this mathematical operation. We notice a pattern emerging out as shown in Figure 7. Further more, the variances in two halves of the dataset which were highly variable does not look too much of a problem as shown in Figure 8. The transformation has greatly aided normality in the data. Figure 10 shows that the normality has improved. It was also noted that the p-value of the Shapiro - Wilk test in the transformed data is 0.016 which still does not indicate a very good fit.

2.2 b

The series y_t and $\log y_t$ are plotted in Figure 6 and Figure 7 respectively. Both these plots show evidence of the cyclical patterns as we saw in global temperature data.

2.3 c

Sample ACF plot of Varve data as shown in Figure 11 clearly indicates

1. Cyclical patterns in the data.
2. Dependency i.e. auto-correlation of higher order lags..

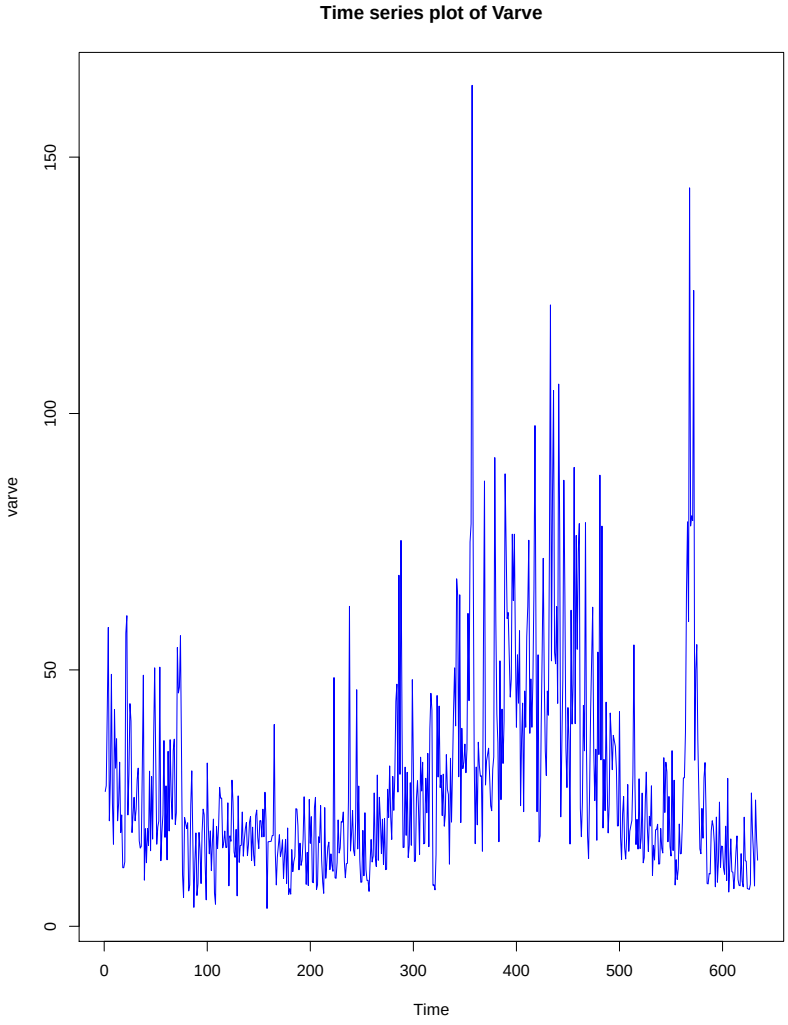

Figure 6: Time Series Plot, Varve

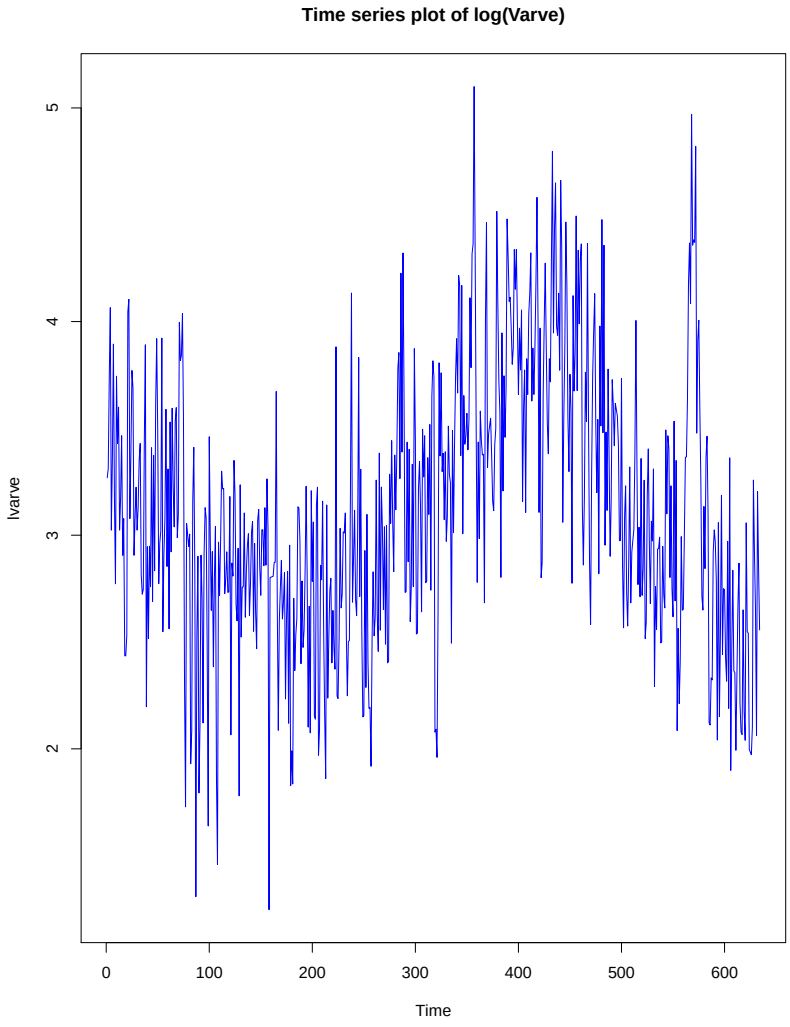

Figure 7: Time Series Plot, log(Varve)

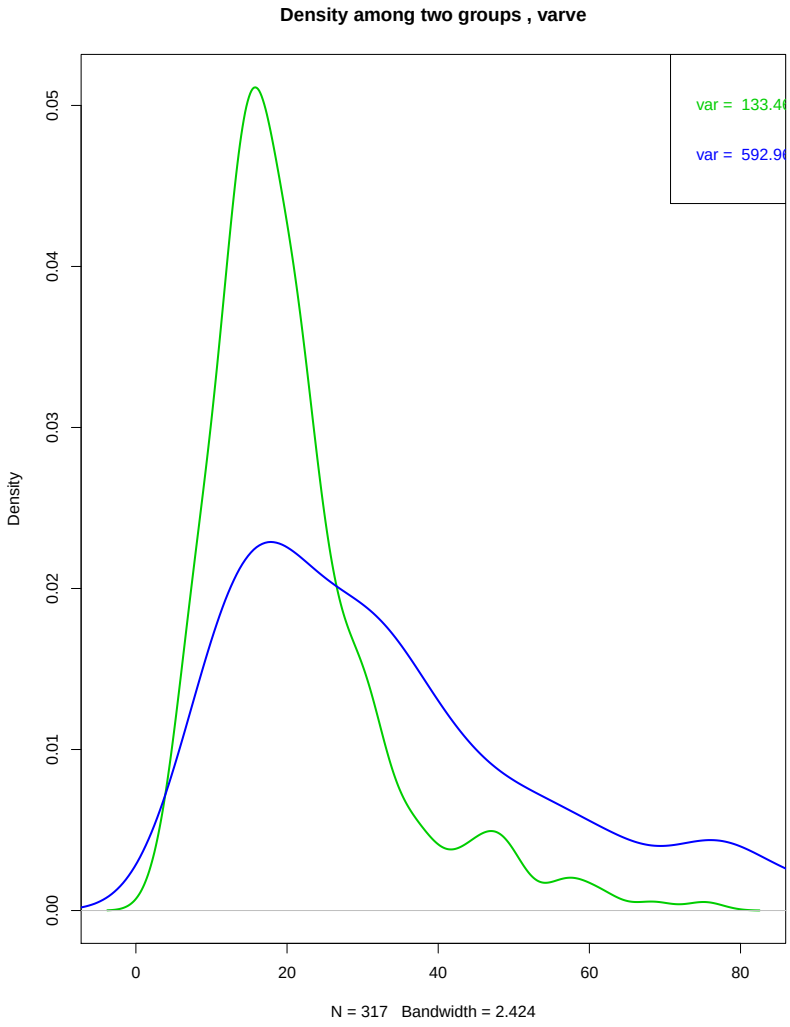

Figure 8: Density Plot, Varve

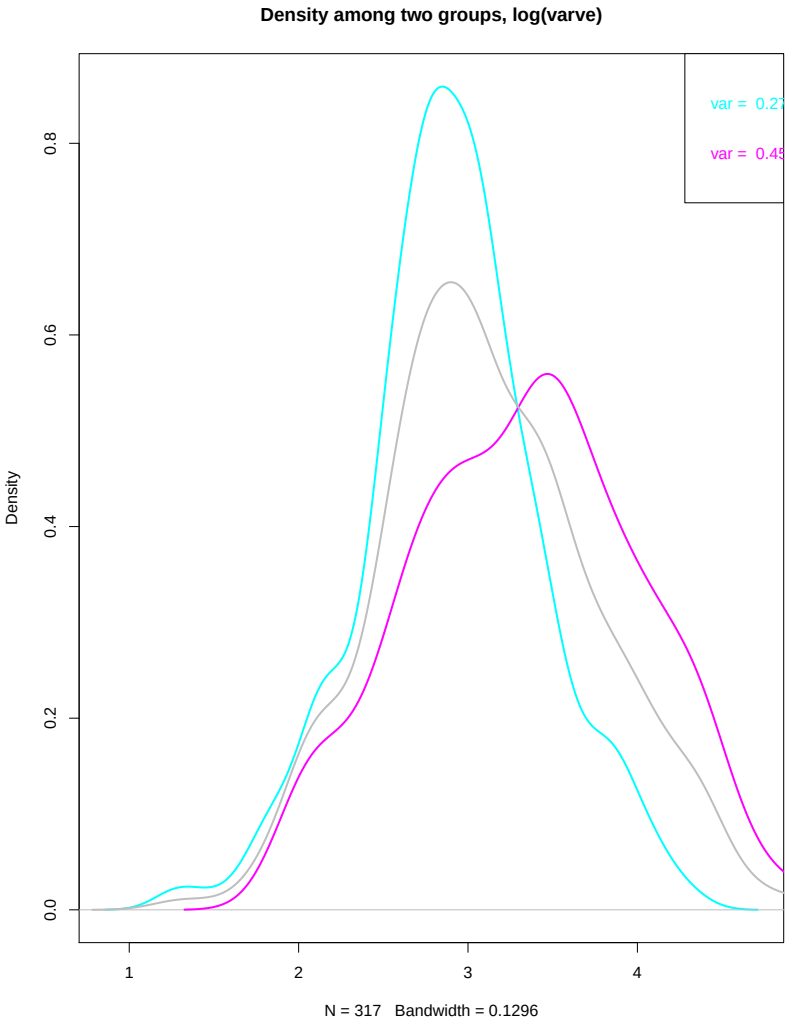

Figure 9: Density Plot, log(Varve)

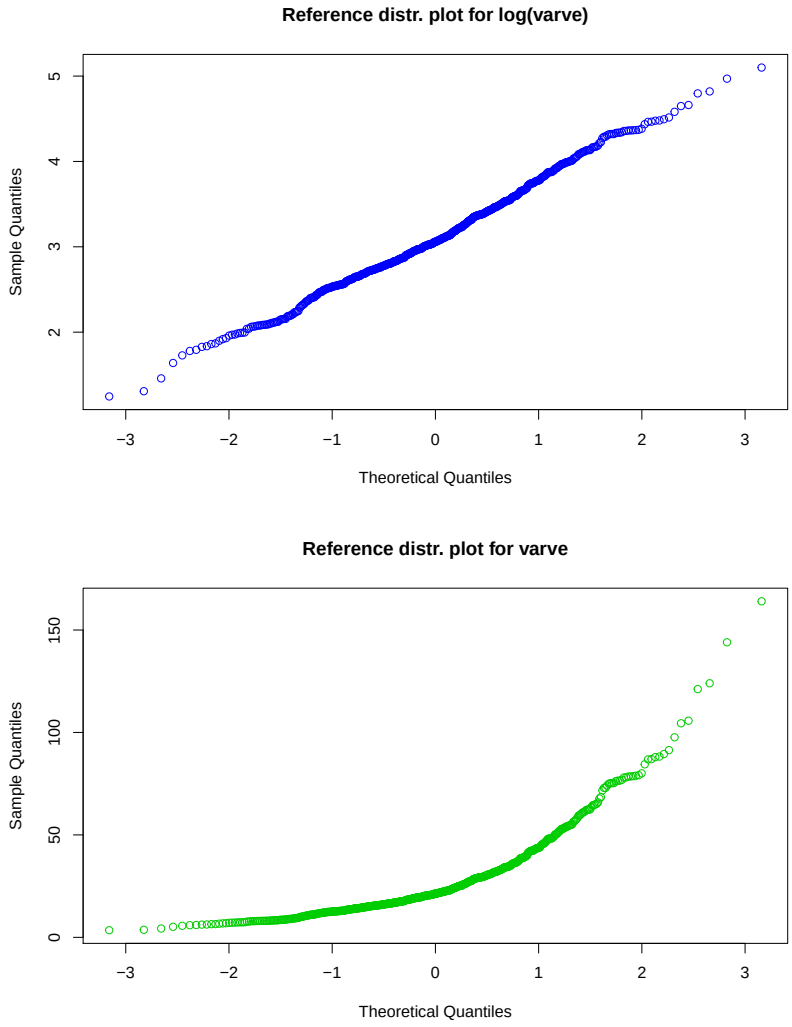

Figure 10: Q-Q Plots before and after transformation

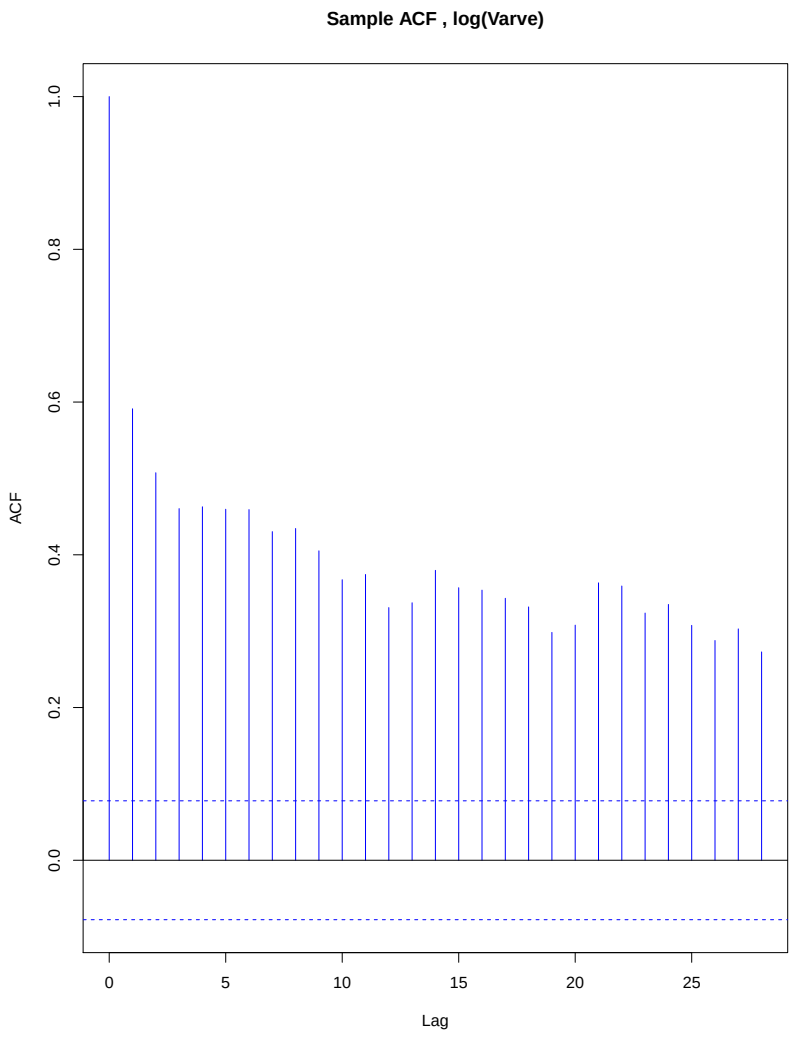

Figure 11: Sample ACF, log(Varve) Data

2.4 d

Sample ACF plot and First order differencing of Varve data as shown in Figure 13 and Figure 12 respectively. Figure 12 also has a smoothed Lowess curve overlayed . Both these figures provide visual support that

1. Any trend or cyclical pattern has been removed from the series.
2. Smoothed curve shows that the expected value from the sample , does not vary with the mean.
3. Sample Acf, does not detect higher order lag.
4. Variance of the data has been controlled and does not vary with either time or lag.
5. The data points look coming from $i.i.d$ distribution with a constant variance

All of the above indicates that the Varve data set has been reduced to a stationary series now.

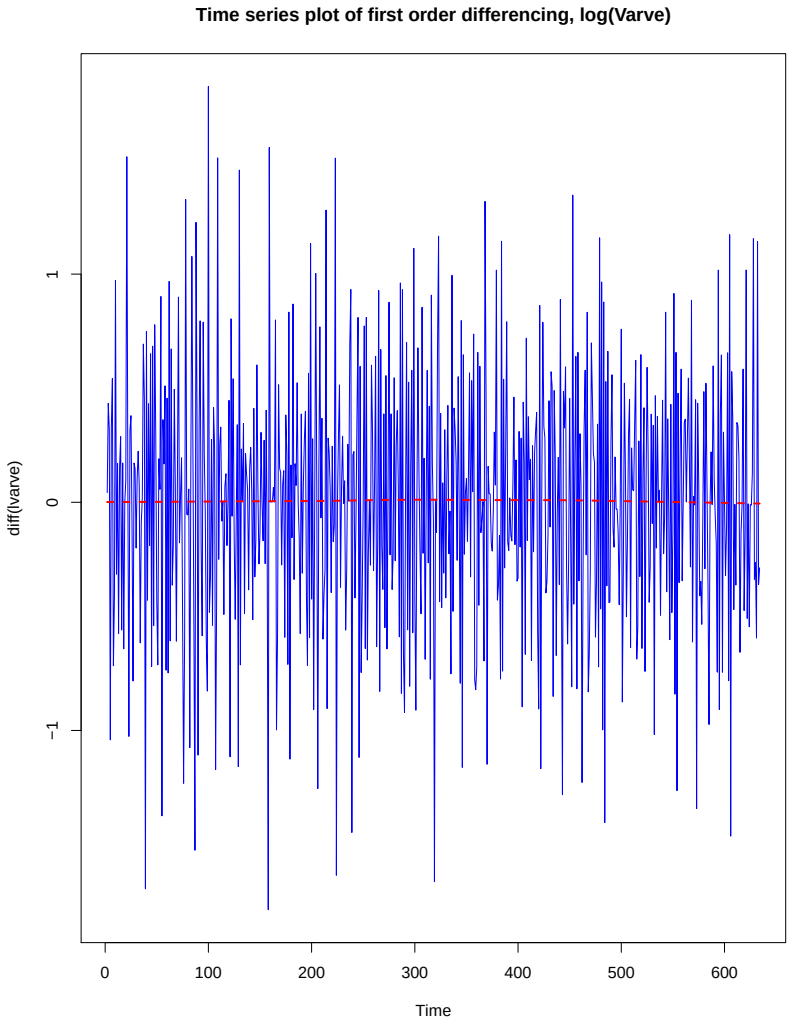

Figure 12: Time Series, 1st Order Differencing, log(Varve) Data

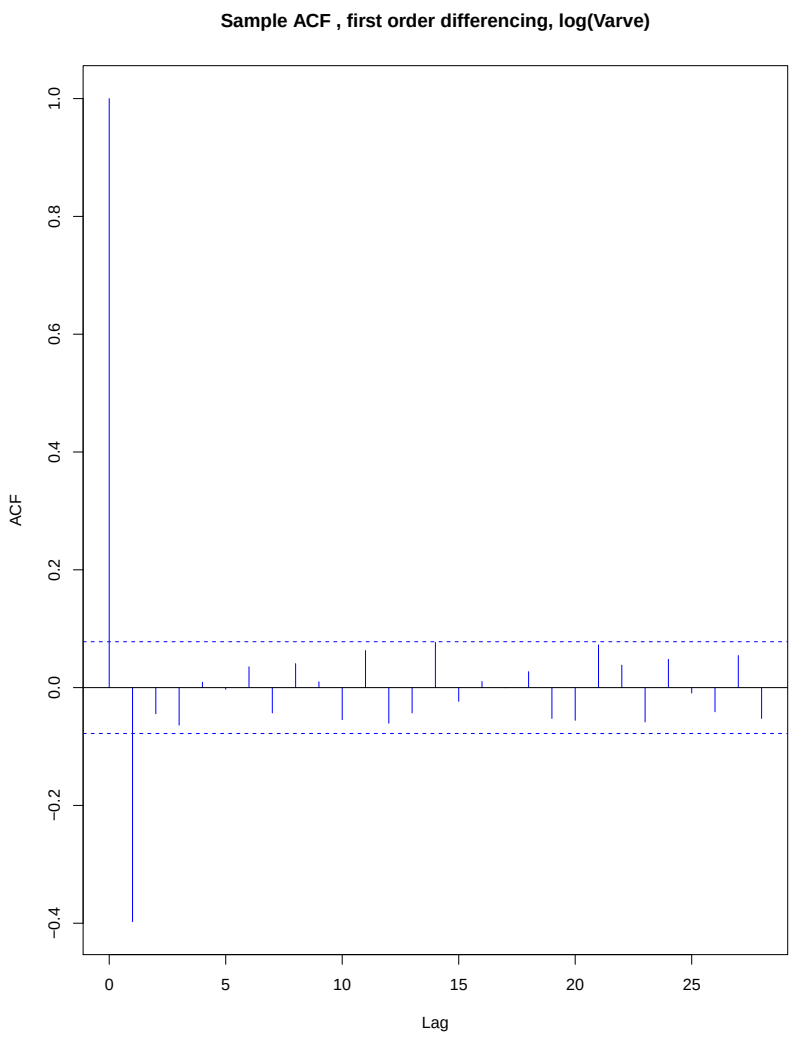

Figure 13: Sample ACF, 1st Order Differencing, log(Varve) Data