Assignment 4

Benjamin Moran

8 September 2016

Introduction

When assessing changes in the Earth's climate over periods of hundreds or thousands of years, climate scientists and geologists often look to layers of sediment or sedimentary rock know as *Rhythmites*. These layers are the result of geological/ecological processes with an easily observable periodicity/regularity and as such can provide insights into any longitudinal changes in the processes that cause their formation.

Varves

A *Varve* is a particular type of *Rhythmite* that forms in the presence of fresh or brackish water. Thus, they often present a record of **deglaciation** in the area where they are observed. By measuring the depth and composition (ratio of silt & sand to clay) of each varve, we can discern a record of the paleoclimactic conditions in that area over a period of thousands of years, or at least for the period during which deglaciation occured.

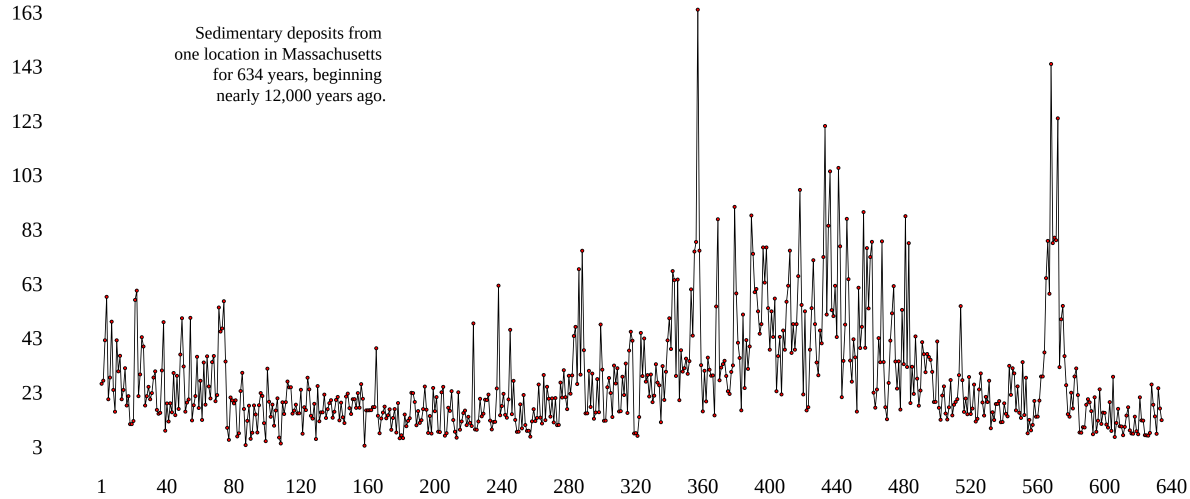
astsa::varve Dataset

The *varve* dataset contained in the *astsa* package for **R**, contains a record of yearly glacial deposits in Massachusetts, New England in a period from approximately 11,834 years ago (the time deglaciation began in the area) to approximately 11,200 years ago (the time deglaciation ended). The time-series data contains a record of the thicknesses of each varve from this location.

Analysis

Plotting the Time-series x_t.

Firstly, we can plot the time-series in its raw form.



Homoscedasticity

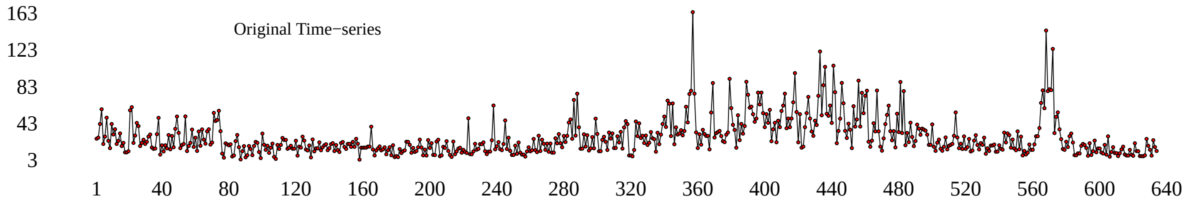
In order for us to perform any kind of analysis on this time-series, we need to confirm that it is **homoscedastic**, or that the variance of the time-series does not change over time. We can calculate the variance of the first half of the time-series and compare it with the variance of the second half to easily check. The variance of the first half ≈ 133.46; the variance of the second half ≈ 594.49. These are significantly different, so we will need to construct a *logarithmic transformation* to ensure that the time-series is *homoscedastic*.

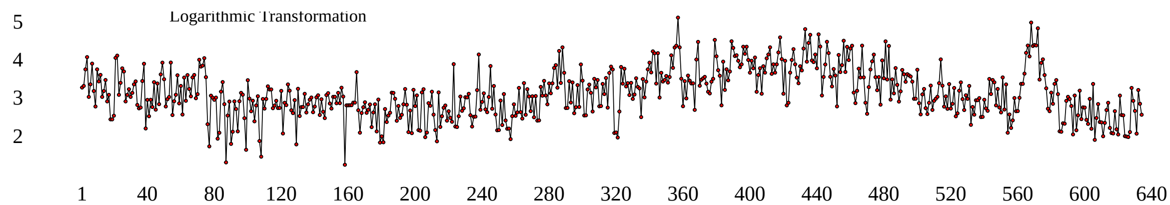
Logarithmic Transformation

We construct a *logarithmic transformation* via the following method: if x_t is a time series that is *heteroscedastic* (its variance changes over time), the the transformation $y_t = log(x_t)$ is a *homoscedastic* time-series.

Plotting the Transformed Time-series y_t.

We can plot the original time-series x_t alongside our transformation $y_t$to observe the changes made by this process.





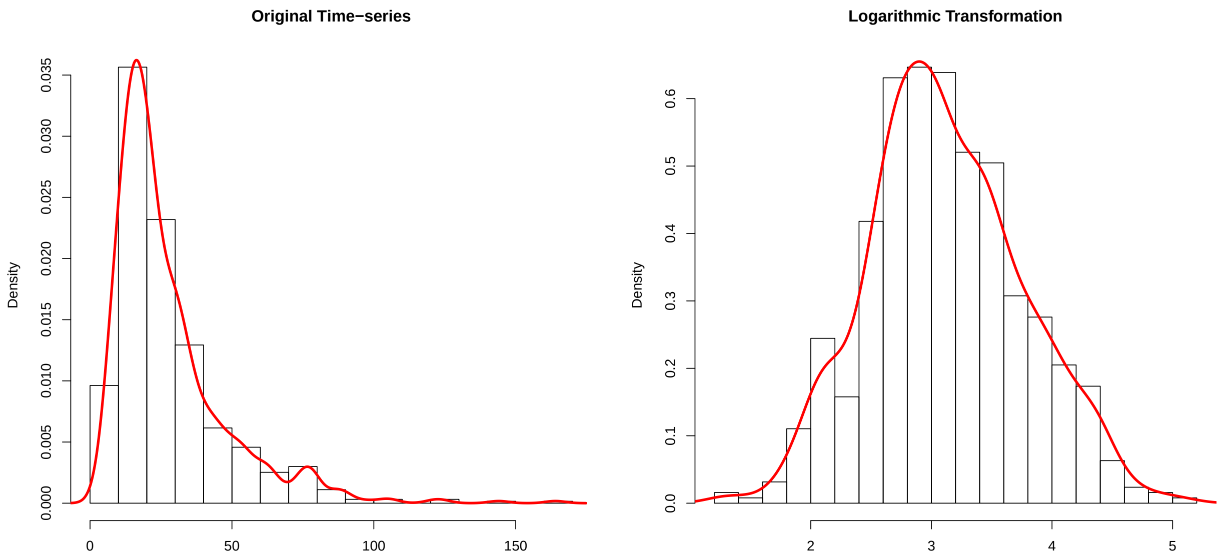
The scale of the data values have greatly reduced in the transformed time-series. The variance of the first half ≈ 0.2707; the variance of the second half ≈ 0.4514. These are an improvement on the previous time-series, but still different, so we can't yet say that the time-series is *homoscedastic*.

Normality

We can further test whether or not the transformed time-series y_t is homoscedastic by performing a **Shapiro-Wilk Test** on the data. If the data is homoscedastic, then the data is also normally-distributed. Thus if the results of running the **Shapiro-Wilk Test** confirm the null-hypothesis, then we can say that the data is homoscedastic and continue with our analysis.

First we plot the histograms of the orginal - x_t - and transformed - y_t - time-series.

Histograms of x_t and y_t.



The original time-series x_t is obviously heavily-skewed and the transformed time-series y_t looks farily skewed as well, suggesting that the transformed time-series is still not homoscedastic. We can confirm this via the Shapiro-Wilk Test.

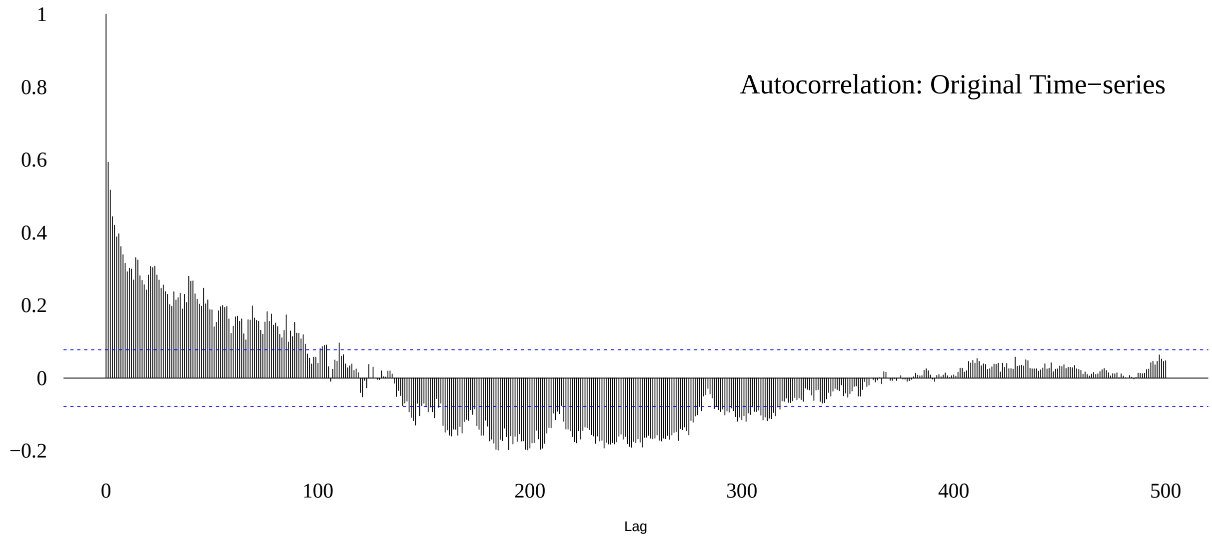
Shapiro-Wilk Test

Running the test on the orginal time-series x_t with a significance level of $\alpha = 0.05$ returns a p-value of ≈ 0 to four decimal places. Therefore, we reject the null-hypothesis that the time-series data is normally distributed, which implies that the original time-series x_t is heteroscedastic. Running the test on the transformed time-series y_t with a significance level of $\alpha = 0.05$ returns a p-value of ≈ 0.0168 to four decimal places. Therefore, we reject the null-hypothesis that the time-series data is normally distributed, which implies that the transformed time-series y_t is heteroscedastic.

Autocorrelation

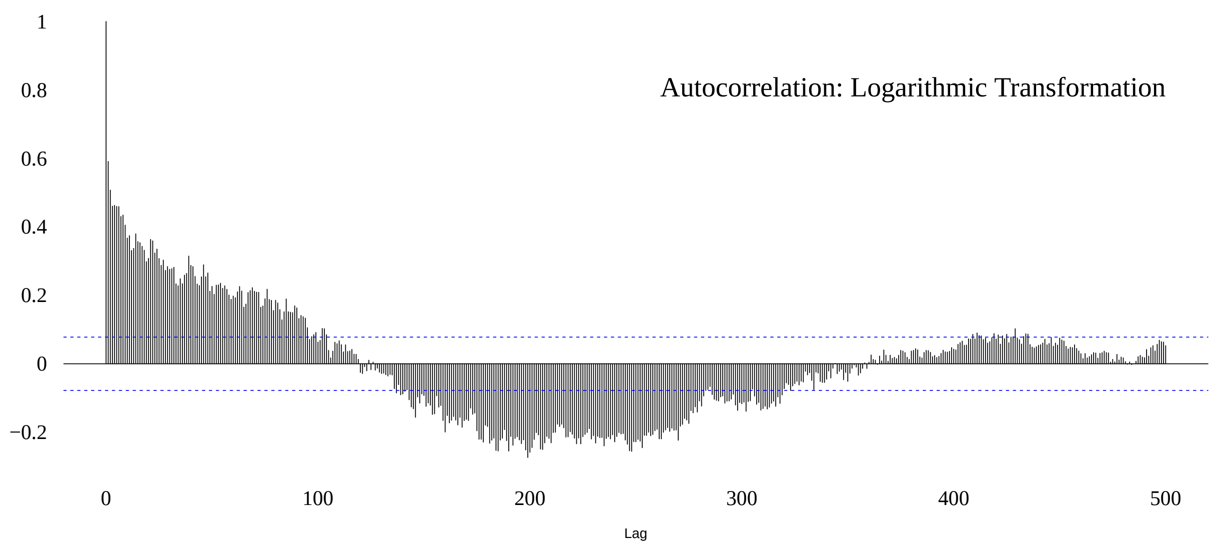
Further proof of the results can be obtained by plotting the Autocorrelation Function (ACF) of each time-series. A time-series is homoscedastic if more than 95% of the values output by the ACF fall within a 95% normal confidence interval: i.e. if this condition holds, then the data is normally distributed.

First, we plot the ACF of the original time-series x_t.



It seems that this data does not meet the conditions for homoscedacity specified above. We can confirm this by evalualting the following expression in R: $mean(abs(acf(x_t, 500)\$acf) < qnorm((1 + 0.95)/2)/sqrt(634)) = 0.4710579 < 0.95 \implies x_t$ is not homoscedastic.

Repeating the same analysis for the transformed time-series y_t.



Again, it seems that this data does not meet the conditions for homoscedacity specified above. We can confirm this by evalualting the following expression in R: $mean(abs(acf(y_t, 500)\$acf) < qnorm((1 + 0.95)/2)/sqrt(634)) = 0.4111776 < 0.95 \implies y_t$ is not homoscedastic.

Differencing

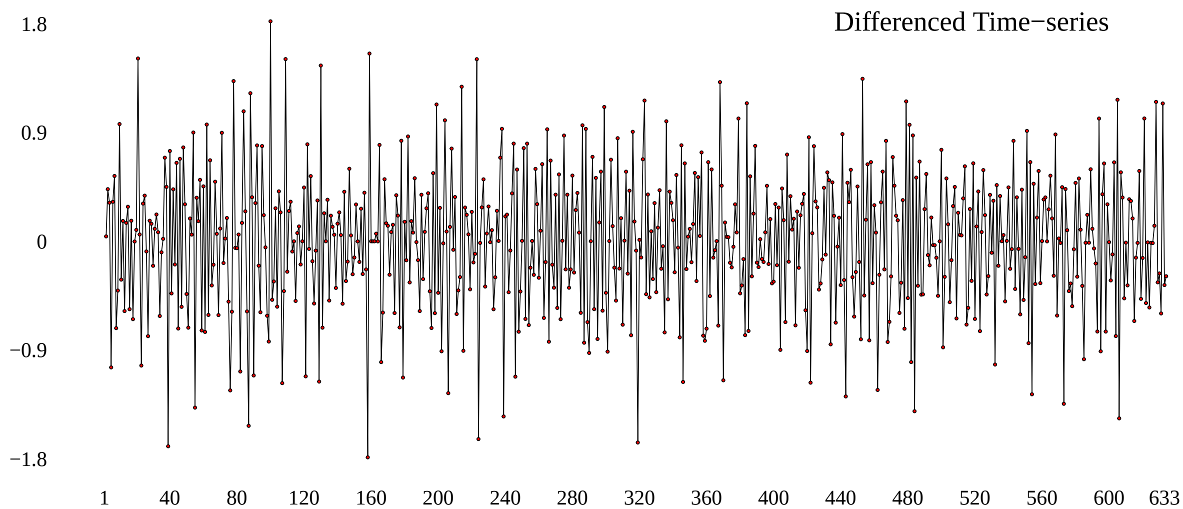
We can create a homoscedastic time-series from the transformed time-series y_t via *Differenceing*. This method involves creating a new time-series u_t using the following process:

$$u_t := \nabla y_t = y_t - y_{t-1}.$$

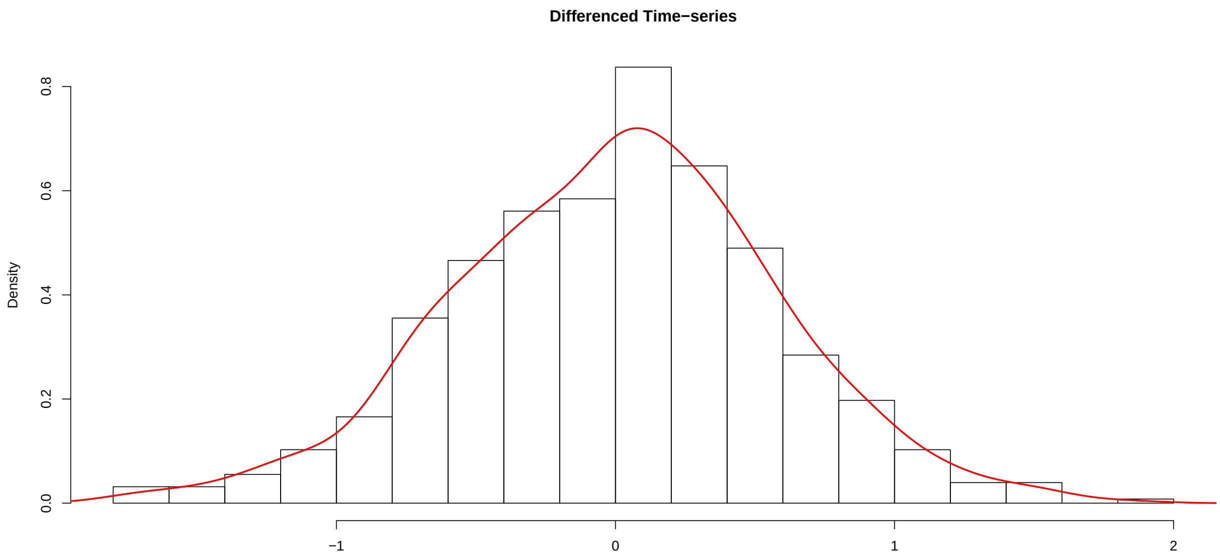
where ∇ is the differencing operator. If we define y_t in the above equation as the transformed time-series form above, we get a new time-series to test.

Plotting the Differenced Time-series u_t.

Firstly, we can plot the differenced time-series u_t:



Histogram of the Differenced Time-series u_t.



This is much more promising. We can confirm this via the Shapiro-Wilk Test.

Shapiro-Wilk Test

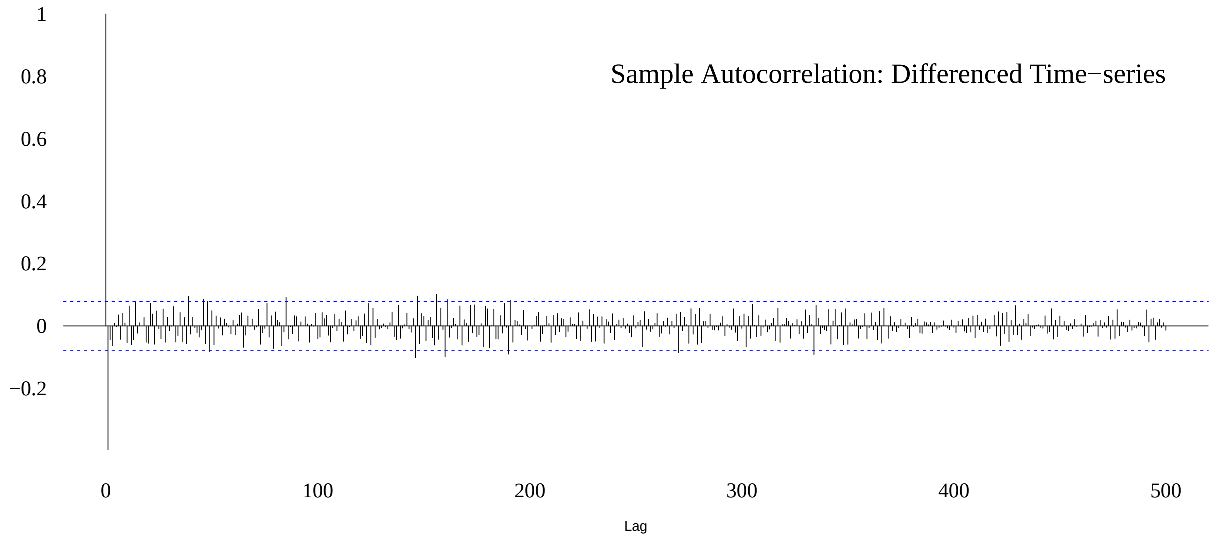
Running the test on the orginal time-series u_t with a significance level of $\alpha = 0.05$ returns a p-value of ≈ 0.6043 to four decimal places. Therefore, we fail to reject the null-hypothesis that the time-series data is normally distributed, which implies that the original time-series u_t is homoscedastic.

Interpretations

All of the transformations we have performed on our time-series data up to this point have been an attempt to arrive at a *stationary* time-series that we can analyse. In order to arrive at this, we have focused on finding a time-series that is *homoscedastic* as a proxy for *stationarity*.
Can you think of a practical interpretation for the time series u_t? *Hint: For* $\mid p \mid$ *close to zero,* $log(1 + p) \approx p$*; let* $p = (x_t - x_{t-1})/x_{t-1}$.

Sample Autocorrelation

Plot the sample ACF of the time series u_t .



We can confirm this by evalualting the following expression in R: $mean(abs(acf(u_t, 500)\$acf) < qnorm((1 + 0.95)/2)/sqrt(634)) = 0.9681 > 0.95 \implies u_t$ is homoscedastic.

Moving Average Model

Based on the sample ACF of the time series u_t, argue that the following model might be reasonable. Consider the moving average model.

$$U_t = \mu + \mathcal{W}_t - \theta \mathcal{W}_{t-1}$$

If $\mathcal{W}_t$ are assumed WN$[0, \sigma^2_{\mathcal{W}}]$, then U_t is a stationary time series. Show that

$$\gamma_U(h) = \begin{cases} \sigma^2_{\mathcal{W}}(1+\theta^2) & \text{for } h = 0 \\ -\theta\sigma^2_{\mathcal{W}} & \text{for } h = \pm 1 \\ 0 & \text{for } \mid h \mid > 1 \end{cases}$$

11. Based on the results of Item 10, use $\widehat{\gamma}_U(0)$ and $\widehat{\gamma}_U(1)$ to derive estimates of θ and $\sigma^2_{\mathcal{W}}$. This is an application of the method of moments from classical statistics, where estimators of the parameters are derived by equating sample moments to theoretical moments.

References: