STAT211 Mandatory Homework 2

Yapi Donatien Achou

February 13, 2019

1 Problem 2.1

let X_t be given by

$$X_t = \beta_0 + \beta_1 t + Z_t \tag{1}$$

where Z_t is an independent random variable with mean 0 and variance σ^2.

1.1 Part A: Prove that X_t is non stationary

$$\begin{aligned} \mu_X(t) &= E[X_t] \\ &= E[\beta_0 + \beta_1 t + Z_t] \\ &= E[\beta_0] + E[\beta_1 t] + E[Z_t] \\ &= \beta_0 + \beta_1 t \end{aligned} \tag{2}$$

Since $\mu_X(t)$ depends on t, we conclude that X_t is non-stationary.

1.2 Part b

let ΔX be given by

$$\Delta X = X_t - X_{t-1} = \beta_1 + Z_t - Z_{t-1} \tag{3}$$

Then

$$\begin{aligned} E[\Delta X] &= E[\beta_1 + Z_t - Z_{t-1}] \\ &= E[\beta_1] + E[Z_t] - E[Z_{t-1}] \\ &= \beta_1 \end{aligned} \tag{4}$$

$$
\begin{aligned}
\gamma_X(t+h,t) &= \mathrm{Cov}(\beta_1 + Z_{t+h} - Z_{t+h-1}, \beta_1 + Z_t - Z_{t-1}) \\
&= \mathrm{Cov}(Z_{t+h} - Z_{t+h-1}, Z_t - Z_{t-1}) \\
&= \mathrm{Cov}(Z_{t+h}, Z_t) - \mathrm{Cov}(Z_{t+h}, Z_{t-1}) - \mathrm{Cov}(Z_{t+h-1}, Z_t) + \mathrm{Cov}(Z_{t+h-1}, Z_{t-1}) \\
&= \sigma_z^2(\delta_{h,0} - \delta_{h,-1} - \delta_{h,1} + \delta_{h,0}) \\
&= \begin{cases} 2\sigma_z^2 & \text{if h=0} \\ -\sigma_z^2 & \text{if h} = \pm\ 1 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
\tag{5}
$$

$E[\Delta X]$ and $\gamma_X(t+h,t)$ are independent of t, therefore ΔX is stationary

2 Part c

If Z_t is replaced by a general process Y_t with mean μ_Y and auto-covariance function $\gamma_Y(h)$.

$$
\begin{aligned}
E[\Delta X] &= E[\beta_1 + Y_t - Y_{t-1}] \\
&= E[\beta_1] + E[Y_t] - E[Y_{t-1}] \\
&= \beta_1 + \mu_Y - \mu_Y \\
&= \beta_1
\end{aligned}
\tag{6}
$$

The mean is independent of t. For the auto-covariance function we use the same computation as in question b.

3 Problem 2.2

3.1 Introduction

In this section we analyse the varve glacial data and perform different statistical transformation of the data. A varve is an annual layer of sediment [wikipedia]

3.2 Part a). Plot

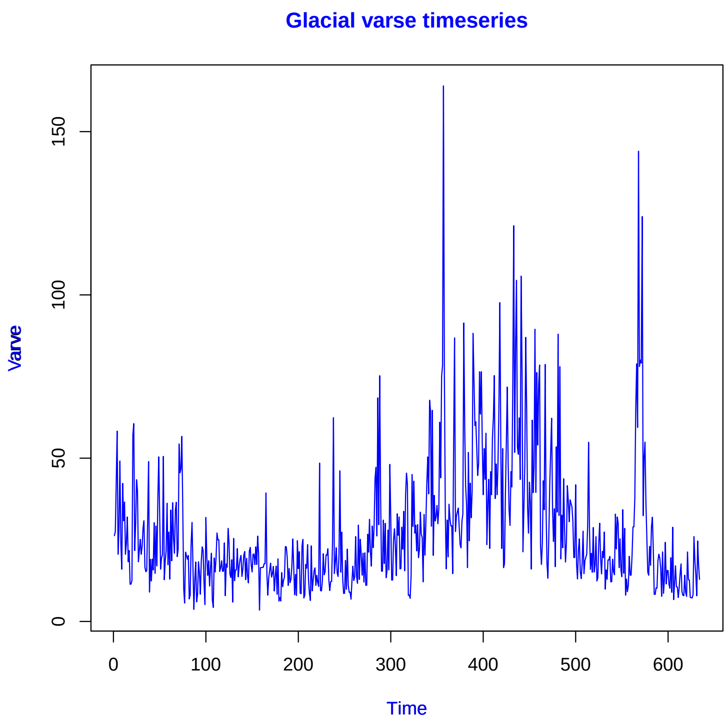

Figure 1: Plot of the glacial varve data

From Figure 1 we see that the varve glacial data exhibits some nonstationarity. To improve the nonstationarity we can use a logarithmic and a differencing transformation.

3.3 Part b). Heteroscedasticity of the varve data

A collection of random variables is heteroscedastic if there are sub-population that have different variability, where the variability can be measured by any statistical dispersion [wikipedia]. We use the variance to show that the varve data exhibits some Heteroscedasticity by computing and comparing the variance of the first and the second half of the data.

Let X_t be the varve time series. The R code bellow computes the variance of the first and second half of X_t.

```
library(astsa)
data(varve)

index <- length(varve)/2
variance_first_half <- var(varve[0:index])
variance_second_half <- var(varve[index:length(varve)])

print(variance_first_half)
>>  133.4574
print(variance_second_half)
>>  592.9645
```

We see that the variance of the second half of the data is approximately 4.44 time larger than the variance of the first half of the data. This shows that the first and the second half of the data have significantly different variability. Therefore the time series X_t exhibits Heteroscedasticity.

Now, let Y_t be the log transformation of X_t

$$Y_t = \log(X_t) \tag{7}$$

```
library(astsa)
data(varve)

index <- length(varve)/2
variance_loga_first_half <- var(log(varve[0:index]))
variance_loga_second_half <- var(log(varve[index:length(varve)]))

print(variance_loga_first_half)
>>  0.2707217
print(variance_loga_second_half)
>>  0.4506843
```

We can observe that after the log transformation the variance of the second half of Y_t is 1.66 time larger than the the first half. This shows that the heteroscedasticity has been significantly reduced.

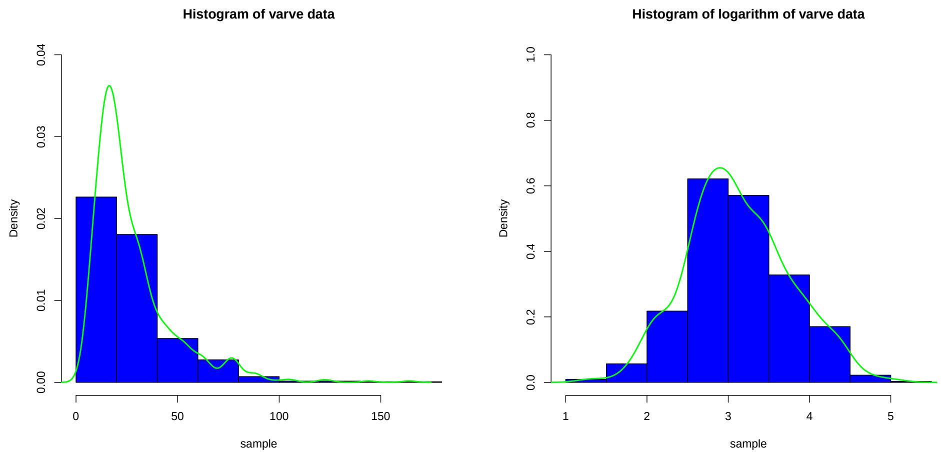

Figure 2: Histogram of varve data (left) and histogram of the log transformation of varve data (right)

From Figure 2 we see that the histogram of the varve data X_t is skewed to the right, but the log transformation $Y_t = \log(X_t)$ is approximately normally distributed.

3.4 Part c). Plot of Y_t

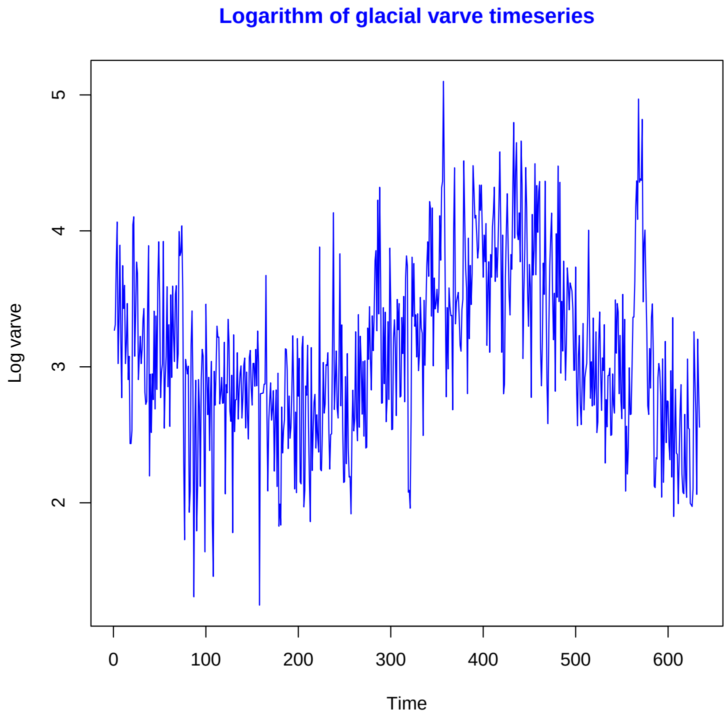

Figure 3: example caption

3.5 Part d). Sample Auto covariance Function of Y_t

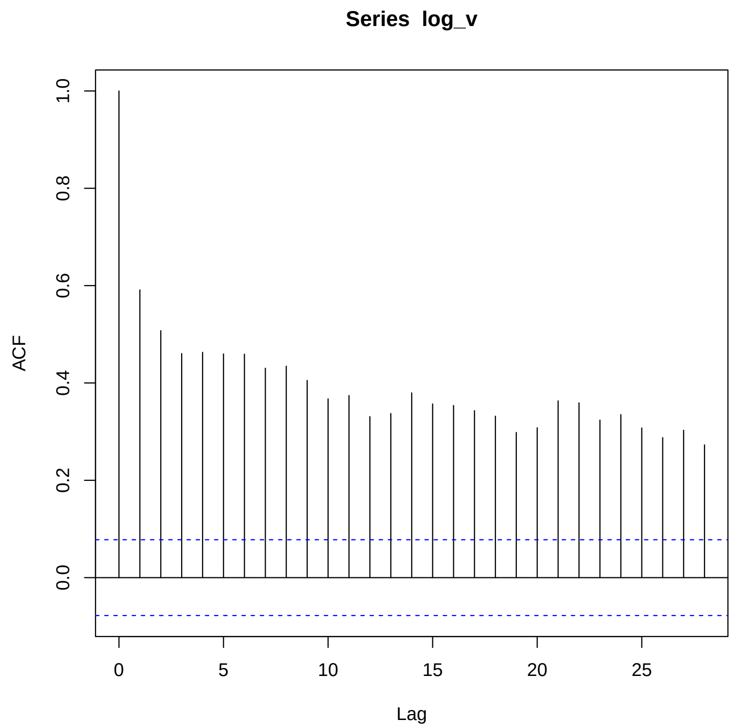

Figure 4: Sample Autocorrelation Function (ACF) of Y_t the log of the varve data and the bound $\pm 1.96\sqrt{n}$ (the dash blue line)

From Figure 4, we see that the sample ACF of Y_t is a decreasing function of lag with local maximum. We also observe some periodicity. Since the sample correlation function of Y_t is slowly decaying with some period, this suggests some trend and seasonality.

4 Part e): Compute the difference $U_t = Y_t - Y_{t-1}$

let U_t be the difference

$$U_t = Y_t - Y_{t-1} = \log(X_t) - \log(X_{t-1}) \tag{8}$$

where X_t is the varve glacial time series.

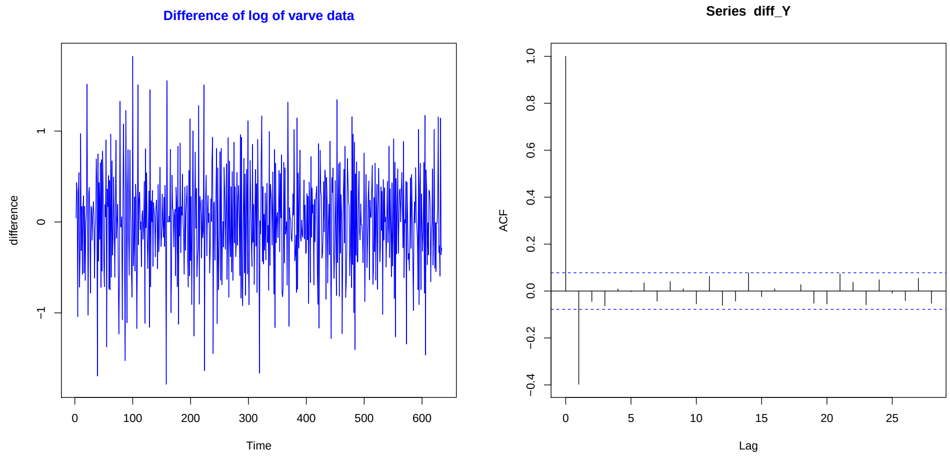

Figure 5: Plot of the first order difference $U_t = Y_t - Y_{t-1}$ (left) and its sample Autocorrelation Function (right)

From Figure 5, we observe that the sample correlation function (right plot) does not exhibits any upwards or downwards trend, showing that the mean is independent of t. There is also no seasonality, suggesting that the variance is the same for all t. Most of the values fall within the bound $\pm 1.96\sqrt{n}$. The correlation at any lag is significantly closer to 0, suggesting that U_{t+h} and U_t are independent. This suggests that the sample correlation function γ_U is independent of t. Therefore, the differencing defined in equation (8) produces a reasonably stationary time series.

The differencing defined in equation (8) can be practically interpreted as a trend remover or trend eliminator.

5 Part f)

Let U_t be modelled by the process

$$U_t = \mu + Z_t + \theta Z_{t-1} \tag{9}$$

This process belonged to the following class of stationary process:

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \tag{10}$$

where

$$\{Z_t\} \sim WN(0, \sigma_z^2) \tag{11}$$

and

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty. \tag{12}$$

By comparing (9) and (10) we see that $\psi_0 = 1$ and $\psi_1 = \theta$. Since U_t is reasonably approximated a stationary time series it can be modelled by (9).

Now let compute $\gamma_u(t+h,t)$ for U_t given by equation (9).

$$\begin{aligned}
\gamma_u(t+h,t) &= \operatorname{Cov}(U_{t+h}, U_t) \\
&= \operatorname{Cov}(\mu + Z_{t+h} + \theta Z_{t+h-1}, \mu + Z_t + \theta Z_{t-1}) \\
&= \operatorname{Cov}(Z_{t+h} + \theta Z_{t+h-1}, Z_t + \theta Z_{t-1}) \\
&= \operatorname{Cov}(Z_{t+h}, Z_t) + \theta \operatorname{Cov}(Z_{t+h}, Z_{t-1}) + \theta \operatorname{Cov}(Z_{t+h-1}, Z_t) + \theta^2 \operatorname{Cov}(Z_{t+h-1}, Z_{t-t}) \\
&= \sigma_z^2 \delta_{h,0} + \theta \sigma_z^2 \delta_{h,-1} + \theta \sigma_z^2 \delta_{h,1} + \theta^2 \sigma_z^2 \delta_{h,0} \\
&= \sigma_z^2 (\delta_{h,0} + \theta \delta_{h,-1} + \theta \delta_{h,1} + \theta^2 \delta_{h,0})
\end{aligned}$$

- if h = 0, then

$$\begin{aligned}
\gamma_u(t+h,t) &= \sigma_z^2(\delta_{0,0} + \theta\delta_{0,-1} + \theta\delta_{0,1} + \theta^2\delta_{0,0}) \\
&= \sigma_z^2(1 + 0\theta + 0\theta + 1\theta^2) \\
&= \sigma_z^2(1+\theta^2)
\end{aligned} \tag{13}$$

- if h = 1, then

$$\begin{aligned}
\gamma_u(t+h,t) &= \sigma_z^2(\delta_{1,0} + \theta\delta_{1,-1} + \theta\delta_{1,1} + \theta^2\delta_{1,0}) \\
&= \sigma_z^2(0 + 0\theta + 1\theta + 0\theta^2) \\
&= \sigma_z^2\theta
\end{aligned} \tag{14}$$

- if h = -1, then

$$\begin{aligned}
\gamma_u(h) &= \sigma_z^2(\delta_{-1,0} + \theta\delta_{-1,-1} + \theta\delta_{-1,1} + \theta^2\delta_{-1,0}) \\
&= \sigma_z^2(0 + 1\theta + 0\theta + 0\theta^2) \\
&= \sigma_z^2\theta
\end{aligned} \tag{15}$$

- if h = a, where $a \neq \pm 1$ and $a \neq 0$

$$\begin{aligned}
\gamma_u(t+h,t) &= \sigma_z^2(\delta_{a,0} + \theta\delta_{a,-1} + \theta\delta_{a,1} + \theta^2\delta_{a,0}) \\
&= \sigma_z^2(0 + 0\theta + 0\theta + 0\theta^2) \\
&= 0
\end{aligned} \tag{16}$$

Therefore

$$\gamma_u(t+h,t) = \begin{cases} \sigma_z^2(1+\theta^2) & \text{if h=0} \\ \theta\sigma_z^2 & \text{if h} = \pm 1 \\ 0 & \text{otherwise} \end{cases}$$

6 Part g

$$\hat{\rho}_u(1) = \frac{\hat{\gamma}_u(1)}{\hat{\gamma}_u(0)} = \frac{\theta}{1+\theta^2} \tag{17}$$

From R code we have

```
library(astsa)
data(varve)
Y <- log(varve)
U <- diff(Y,lag=1, differences=1)
emp_auto_corr_rho <- acf(U,type = "correlation")
emp_auo_variance_gamma <- acf(U, type = "covariance")
print(emp_auto_corr_rho[1])

>> -0.397
```

We have

$$\hat{\rho}_u(1) = \frac{\hat{\gamma}_u(1)}{\hat{\gamma}_u(0)} = \frac{\theta}{1+\theta^2} = -0.397 \tag{18}$$

From which we get

$$0.397 + \theta + 0.397\theta^2 = 0 \tag{19}$$

with

$$\theta_1 \approx -0.49 \quad \theta_2 \approx -2 \tag{20}$$

Now

$$\begin{aligned} \operatorname{Var}(U) &= \operatorname{Cov}(U,U) \\ &= \operatorname{Cov}(\mu + Z_t + \theta Z_{t-1}, \mu + Z_t + \theta Z_{t-1}) \\ &= \operatorname{Cov}(Z_t, Z_t) + \theta \operatorname{Cov}(Z_t, Z_{t-1}) + \theta \operatorname{Cov}(Z_{t-1}, Z_t) + \theta^2 \operatorname{Cov}(Z_{t-1}, Z_{t-1}) \\ &= \sigma_z^2(1+2\theta+\theta^2) \end{aligned} \tag{21}$$

From the data the sample variance is

```
library(astsa)
data(varve)
Y <- log(varve)
U <- diff(Y,lag=1, differences=1)
variance_U <- var(U)
print(variance_U)
>> 0.3322131
```

and

$$\sigma_z^2 = \frac{\mathrm{Var}(U)}{(1+2\theta+\theta^2)} \tag{22}$$

from which we get

$$\sigma_z^2 \approx 1.32726 \quad or \quad \sigma_z^2 \approx 0.3322131 \tag{23}$$

The second one is the variance of U, so $\sigma_z^2 \approx 1.32726$ and $\theta \approx -0.49$?

7 Problem 2.3

$$X_t = Z_t + \theta Z_{t-1} \tag{24}$$

We know that

$$\rho_{X_{t+h},X_t} = \frac{\mathrm{Cov}(X_{t+h}, X_t)}{\sigma_{X_{t+h}}\sigma_{X_t}} \tag{25}$$

Also

$$\mathrm{Cov}(X_{t+h}, X_t) = \gamma_u(h) = \begin{cases} \sigma_z^2(1+\theta^2) & \text{if h=0} \\ \theta\sigma_z^2 & \text{if h} = \pm\ 1 \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

Therefor

$$X_t = Z_t + \theta Z_{t-1} \tag{27}$$

We know that

$$\begin{aligned} \rho_{X_{t+h},X_t} &= \frac{\mathrm{Cov}(X_{t+h}, X_t)}{\sigma_{X_{t+h}}\sigma_{X_t}} \\ &= \begin{cases} \frac{\sigma_z^2(1+\theta^2)}{\sigma_z^2} & \text{if h=0} \\ \frac{\theta\sigma_z^2}{\sigma_z^2} & \text{if h} = \pm\ 1 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} (1+\theta^2) & \text{if h=0} \\ \theta & \text{if h} = \pm\ 1 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{28}$$

If $h = 0$, (X,Y,Z) = (1,3,2)

$$\begin{aligned} \rho_{XY|Z} &= \frac{(1+\theta^2) - (1+\theta^2)^2}{1-(1+\theta^2)^2} \\ &= \frac{1+\theta^2}{2+\theta^2} \end{aligned} \tag{29}$$

If $h = \pm 1$, (X,Y,Z) = (1,3,2)

$$\begin{aligned} \rho_{XY|Z} &= \frac{\theta-\theta^2}{1-\theta^2} \\ &= \frac{\theta}{1+\theta} \end{aligned} \tag{30}$$

Now we have

$$\rho_{XY|Z} = \begin{cases} \frac{1+\theta^2}{2+\theta^2} & \text{if h=0} \\ \frac{\theta}{1+\theta} & \text{if h} = \pm 1 \\ 0 & \text{otherwise} \end{cases} \tag{31}$$

We can says that

$$\rho_{XY|Z} \leq \rho_{XY} \tag{32}$$

8 All R code Code

```
options( warn = -1 )

library(astsa)
data(varve)

# plot data
plot_data <- function(data){
  plot(varve,col="blue")
  title(main="Logarithm␣of␣glacial␣varve␣timeseries", col.main="blue")
  title(xlab="Time", col.lab="blue")
  title(ylab="Varve", col.lab="blue")
}

#histogram plot

histogram <- function(sample){
  hist(sample,main="Histogram␣of␣logarithm␣of␣varve␣data", col="blue", prob=TRUE,yl

  lines(density(sample),lwd=2,col="green")
}

# compute variance of a sample
get_variance <- function(sample){
  sample_variance <- var(sample)
  return(sample_variance)
}

# compute the logarithm of a sample
get_log <- function(sample){
  log_sample <- log(sample)
  plot(log_sample,col="blue",xlab="Time",ylab="Log␣varve", col.lab="blue")
  title(main="Logarithm␣of␣glacial␣varve␣timeseries", col.main="blue")
}
```

```r
plot_difference <- function(sample){
  difference <- diff(sample,lag=1, differences=1)
  plot(difference,col="blue")
  title(main="Difference␣of␣log␣of␣varve␣data", col.main="blue")
  #title(xlab="Time", col.lab="blue")
  #title(ylab="Varve", col.lab="blue")

}



Y <- log(varve)
U <- diff(Y,lag=1, differences=1)
#emp_auto_corr_rho <- acf(U,type = "correlation")
#emp_auo_variance_gamma <- acf(U, type = "covariance")
#print(emp_auo_variance_gamma[1])
#print(emp_auo_variance_gamma[0])

x1 <- var(U)/(1-2*0.49+0.49*0.47)
x2 <- var(U)/(1-2*2+2*2)

print(x1)

print(x2)

print(var(U))
```