

ORIE5550 - HW1 - Ic2234

Luis Alonso Cendra Villalobos (lc2234)

2024-02-01

install.packages("tinytex")

Problem 1

Find 2 univariate time series in different fields online and do the following for each of the two series:

(a) Indicate the exact source for the time series data.

Source of data1: World Bank. URL: <https://data.worldbank.org/indicator/SE.PRM.CMPT.FE.ZS?locations=1W&start=1973&view=chart>

Source of data2: Yahoo! Finance. URL: <https://finance.yahoo.com/quote/AAPL/history?period1=154872000&period2=1705486400&interval=1mo&filter=history&frequency=1mo&includeAdjustedClose=true>

(b) Output the first 20 elements x1,...,x20 of the series in R.

```
#Set working directory
setwd("~/Users/Alonso/OneDrive - Cornell University/Cornell University/Spring 2024/ORIE 5550/ORIE5550_Homework1")
#extract data from csv
data1 <- read.csv("ORIE5550_DataHW1.csv")
data1$Year <- as.Date(data1$Year)
data2 <- read.csv("ORIE5550_Data2HW1.csv")
data2$date <- as.Date(data2$date)

#transform data to Time Series
time_series_data1 <- ts(data1$value, start = min(data1$Year), frequency = 1)
time_series_data2 <- ts(data2$value, start = min(data2$date), frequency = 12)

# Output the first 20 values
first_20_values1 <- head(time_series_data1, 20)
print(first_20_values1)

## [1] 70.67788 69.07636 68.12719 68.04817 71.75885 73.93382 73.02761 73.40257
## [9] 73.72074 73.15771 75.41759 75.92915 76.64542 76.75523 77.51466 77.79708
## [17] 77.87164 77.33327 77.51179 76.97076

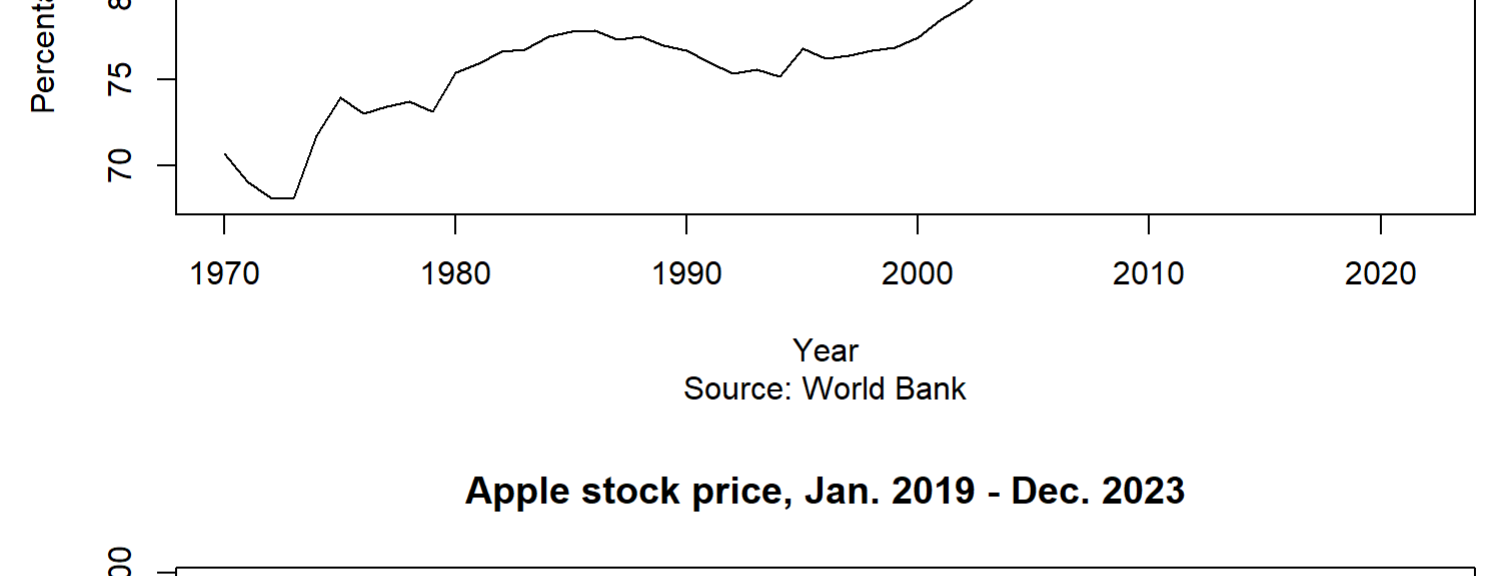
first_20_values2 <- head(time_series_data2, 20)
print(first_20_values2)

## [1] 41.54781 45.77450 48.35781 42.18869 47.87880 51.53047 50.49627
## [8] 54.30649 60.48013 64.89605 71.52080 75.30364 66.57903 62.08135
## [15] 71.72715 77.62961 89.38191 104.04848 126.35438 113.60417
```

(c) Produce a time plot of the series in R, after transforming the series into a time series object.

```
par(mfrow=c(2,1))
plot.ts(time_series_data1, ylab="Percentage(%)", xlab="Year",
        main = "Percentage females with complete primary education (World), 1970 - 2020",
        sub = "Source: World Bank")

plot.ts(time_series_data2, ylab="Price", xlab="Date",
        main = "Apple stock price, Jan. 2019 - Dec. 2023",
        sub = "Source: Yahoo! Finance")
```



(d) Discuss briefly possible objectives for analyzing the time series.

The percentage of females with complete primary education can be analyzed to study the factors that are involved in the literacy rate of women worldwide, we can regress the series against others like investment in education, inclusion of women to the education system, gender equality policies/decisions, among others. The Apple stock price can be examined to determine factors that can forecast the stock price in the future with the objective of making investment decisions, for example.

Problem 2

(a) Produce two different realizations x_t , $t = 1, \dots, T$, of the model of length $T = 100$; Include the R code;

```
# Set the length of the time series
length_of_series <- 100

# Initialize vectors to store the time series values
X_1 <- numeric(length_of_series)
Y_1 <- numeric(length_of_series)
X_2 <- numeric(length_of_series)
Y_2 <- numeric(length_of_series)
Z_1 <- numeric(length_of_series)
Z_2 <- numeric(length_of_series)

# Set the initial value of Z
Z_1[1] <- sample(c(2, -1), size = 1, prob = c(1/3, 2/3)) #realization 1 of Z[1]
Z_2[1] <- sample(c(2, -1), size = 1, prob = c(1/3, 2/3)) #realization 2 of Z[1]

# Generate the time series realizations for Z_t, X_t, and X_t * X_{t-1}
for (t in 2:length_of_series) {
  Z_1[t] <- sample(c(2, -1), size = 1, prob = c(1/3, 2/3))
  Z_2[t] <- sample(c(2, -1), size = 1, prob = c(1/3, 2/3))

  # Realization 1 of X_t and Y = X_t * X_{t-1}
  X_1[t] <- 2 * Z_1[t] + Z_1[t - 1]
  Y_1[t] <- X_1[t] * X_1[t - 1]

  # Realization 2 of X_t and Y = X_t * X_{t-1}
  X_2[t] <- 2 * Z_2[t] + Z_2[t - 1]
  Y_2[t] <- X_2[t] * X_2[t - 1]
}

# Calculate the mean of Z
mean_Z_1 <- mean(Z_1) #mean of realization 1
mean_Z_2 <- mean(Z_2) #mean of realization 2

# Calculate the mean of Z^2
mean_Z_1_squared <- mean(Z_1^2) #mean of realization 1 squared
mean_Z_2_squared <- mean(Z_2^2) #mean of realization 2 squared

# Display the results
print(paste("Mean of Z_1:", mean_Z_1))

## [1] "Mean of Z_1: -0.04"

print(paste("Mean of Z_1^2:", mean_Z_1_squared))

## [1] "Mean of Z_1^2: 1.96"

print(paste("Mean of Z_2:", mean_Z_2))

## [1] "Mean of Z_2: -0.25"

print(paste("Mean of Z_2^2:", mean_Z_2_squared))

## [1] "Mean of Z_2^2: 1.75"
```

(b) Compute theoretically $E(x_t)$ and $E(x_t^2)$; Compare these quantities with the two realizations above; Include the R code.

```
# Calculate the mean of X_t
mean_X_1 <- mean(X_1)
mean_X_2 <- mean(X_2)

# Display the results
print(paste("Mean of X_1:", mean_X_1))

## [1] "Mean of X_1: -0.09"

print(paste("Mean of X_2:", mean_X_2))

## [1] "Mean of X_2: -0.72"

# Calculate the mean of X_t squared
mean_X_1_squared <- mean(X_1^2)
mean_X_2_squared <- mean(X_2^2)

# Display the results
print(paste("Mean of X_1^2:", mean_X_1_squared))

## [1] "Mean of X_1^2: 10.71"

print(paste("Mean of X_2^2:", mean_X_2_squared))

## [1] "Mean of X_2^2: 9.18"
```

(c) Compute theoretically $E(X_t X_{t-1})$ and $\text{Corr}(X_t, X_{t-1})$; Compare these quantities with the two realizations above; Include the R code;

```
# Calculate the mean of Y = X_t * X_{t-1}
mean_Y_1 <- mean(Y_1)
mean_Y_2 <- mean(Y_2)

# Display the results
print("Let Y = X_t * X_{t-1}, then:")

## [1] "Let Y = X_t * X_{t-1}, then:"

print(paste("Mean of Y_1:", mean_Y_1))

## [1] "Mean of Y_1: 5.13"

print(paste("Mean of Y_2:", mean_Y_2))

## [1] "Mean of Y_2: 4.23"

# Correlations
V_11 <- X_1[2:length_of_series]
V_12 <- X_1[1:(length_of_series - 1)]

V_21 <- X_2[2:length_of_series]
V_22 <- X_2[1:(length_of_series - 1)]

# Calculate the correlation between V_1 and V_2
correlation_V1_V2_1 <- cor(V_11, V_12)
correlation_V1_V2_2 <- cor(V_21, V_22)

# Display the correlation
print(paste("Correlation between V_1 and V_2 realization 1:", correlation_V1_V2_1))

## [1] "Correlation between V_1 and V_2 realization 1: 0.48977145322455"

print(paste("Correlation between V_1 and V_2 realization 2:", correlation_V1_V2_2))

## [1] "Correlation between V_1 and V_2 realization 2: 0.43186766739816"
```

(d) What are the possible values of X_t and the probabilities that X_t takes these values? Compare your answers from the two realizations above

See handwritten notes.

Problem 3

Come up with your own time series model and repeat parts (a)-(c) of Problem 2 for the model. Your model should have at least a trend or a periodic component, and incorporate IID noise in some way

See Handwritten notes. We assume $T=100$, $\alpha = 2$, variance = 2.

```
# Set the length of the time series
length_of_series <- 100

# Initialize vectors to store the time series values
X_1 <- numeric(length_of_series)
Y_1 <- numeric(length_of_series)
X_2 <- numeric(length_of_series)
Y_2 <- numeric(length_of_series)
Z_1 <- numeric(length_of_series)
Z_2 <- numeric(length_of_series)

# Create a time series Z_t with elements distributed normally
Z_1 <- rnorm(length_of_series, mean = 0, sd = sqrt(2))
Z_2 <- rnorm(length_of_series, mean = 0, sd = sqrt(2))

X_1[1] <- 2 + Z_1[1]
X_2[1] <- 2 + Z_2[1]

# Generate the time series realizations for Z_t, X_t, and X_t * X_{t-1}
for (t in 2:length_of_series) {
  # Realization 1 of X_t and Y = X_t * X_{t-1}
  X_1[t] <- 2 * t + Z_1[t]
  Y_1[t] <- X_1[t] * X_1[t - 1]

  # Realization 2 of X_t and Y = X_t * X_{t-1}
  X_2[t] <- 2 * t + Z_2[t]
  Y_2[t] <- X_2[t] * X_2[t - 1]
}

# Calculate the mean of X_t
mean_X_1 <- mean(X_1)
mean_X_2 <- mean(X_2)

# Display the results
print(paste("Mean of X_1:", mean_X_1))

## [1] "Mean of X_1: 101.137590232438"

print(paste("Mean of X_2:", mean_X_2))

## [1] "Mean of X_2: 101.015028634881"

# Calculate the mean of X_t squared
mean_X_1_squared <- mean(X_1^2)
mean_X_2_squared <- mean(X_2^2)

# Display the results
print(paste("Mean of X_1^2:", mean_X_1_squared))

## [1] "Mean of X_1^2: 13553.4612728173"

print(paste("Mean of X_2^2:", mean_X_2_squared))

## [1] "Mean of X_2^2: 13552.037203831"

# Calculate the mean of Y = X_t * X_{t-1}
mean_Y_1 <- mean(Y_1)
mean_Y_2 <- mean(Y_2)

# Display the results
print("Let Y = X_t * X_{t-1}, then:")

## [1] "Let Y = X_t * X_{t-1}, then:"

print(paste("Mean of Y_1:", mean_Y_1))

## [1] "Mean of Y_1: 13348.6767750323"

print(paste("Mean of Y_2:", mean_Y_2))

## [1] "Mean of Y_2: 13348.6614269418"

# Correlations
V_11 <- X_1[2:length_of_series]
V_12 <- X_1[1:(length_of_series - 1)]
V_21 <- X_2[2:length_of_series]
V_22 <- X_2[1:(length_of_series - 1)]

# Calculate the correlation between V_1 and V_2
correlation_V1_V2_1 <- cor(V_11, V_12)
correlation_V1_V2_2 <- cor(V_21, V_22)

# Display the correlation
print(paste("Correlation between V_1 and V_2 realization 1:", correlation_V1_V2_1))

## [1] "Correlation between V_1 and V_2 realization 1: 0.99944488187673"

print(paste("Correlation between V_1 and V_2 realization 2:", correlation_V1_V2_2))

## [1] "Correlation between V_1 and V_2 realization 2: 0.999159932019141"
```

Problem 4

(a)

See Handwritten notes.

(b) Produce two different realizations x_t , $t = 1, \dots, T$, of the model of length $T = 100$; Include the R code;

```
# Set the parameters
mean_value <- 1
variance_value <- 4
TT <- 100
num_simulations <- 10000

# Initialize array to store values at T=100 from Monte Carlo
X_values_at_T_1 <- numeric(num_simulations)
X_values_at_T_2 <- numeric(num_simulations)
Y_values_at_T_1 <- numeric(num_simulations)
Y_values_at_T_2 <- numeric(num_simulations)

# Perform simulations
for (sim in 1:num_simulations) {
  # Generate random variable time series Z_t white noise Normal(1,4)
  Z_1_t <- rnorm(TT, mean = mean_value, sd = sqrt(variance_value))
  Z_2_t <- rnorm(TT, mean = mean_value, sd = sqrt(variance_value))

  # Initialize X_t with X_0 = 0 for the 2 realizations
  X_1_t <- numeric(TT)
  X_2_t <- numeric(TT)
  Y_1 <- numeric(TT)
  Y_2 <- numeric(TT)
  X_1_t[1] <- 0
  X_2_t[1] <- 0

  # Generate AR(1) model for both realizations
  for (t in 2:TT) {
    X_1_t[t] <- X_1_t[t-1] + Z_1_t[t]
    X_2_t[t] <- X_2_t[t-1] + Z_2_t[t]

    # Realizations of Y = X_t * X_{t-1}
    Y_1[t] <- X_1_t[t] * X_1_t[t - 1]
    Y_2[t] <- X_2_t[t] * X_2_t[t - 1]
  }

  # Store the last element of X_t in values_at_T for both realizations
  X_values_at_T_1[sim] <- X_1_t[TT]
  X_values_at_T_2[sim] <- X_2_t[TT]
  Y_values_at_T_1[sim] <- Y_1[TT]
  Y_values_at_T_2[sim] <- Y_2[TT]
}
```

(c) Compute theoretically $E(x_t)$ and $E(x_t^2)$ at $t = 100$; Write the R code to compute these quantities (Note: Do not plug in the number into the theoretical result).

```
# Calculate the mean of the realizations
X_mean_1 <- mean(X_values_at_T_1) #mean of realization 1
X_mean_2 <- mean(X_values_at_T_2) #mean of realization 2

# Calculate the mean of Z^2
X_mean_1_squared <- mean(X_values_at_T_1^2) #mean of realization 1 squared
X_mean_2_squared <- mean(X_values_at_T_2^2) #mean of realization 2 squared

# Display the results
print(paste("Mean of realization 1 of X:", X_mean_1))

## [1] "Mean of realization 1 of X: 99.095224496095"

print(paste("Mean of realization 1 of X squared:", X_mean_1_squared))

## [1] "Mean of realization 1 of X squared: 10218.7946025257"

print(paste("Mean of realization 2 of X:", X_mean_2))

## [1] "Mean of realization 2 of X: 98.715521359747"

print(paste("Mean of realization 2 of X squared:", X_mean_2_squared))

## [1] "Mean of realization 2 of X squared: 10131.1417835615"
```

(d) Compute theoretically $E(X_t * X_{t-1})$ and $\text{Corr}(X_t, X_{t-1})$ at $t = 100$; Write the R code to compute these quantities (Note: Do not plug in the number into the theoretical result).

```
# Calculate the mean of Y = X_t * X_{t-1}
mean_Y_1 <- mean(Y_values_at_T_1)
mean_Y_2 <- mean(Y_values_at_T_2)

# Display the results
print("Let Y = X_t * X_{t-1}, then:")

## [1] "Let Y = X_t * X_{t-1}, then:"

print(paste("Mean of realization 1 of Y:", mean_Y_1))

## [1] "Mean of realization 1 of Y: 10118.4295853896"

print(paste("Mean of realization 2 of Y:", mean_Y_2))

## [1] "Mean of realization 2 of Y: 10118.4295853896"

# Correlations
V_11 <- X_values_at_T_1[2:length_of_series]
V_12 <- X_values_at_T_1[1:(length_of_series - 1)]
V_21 <- X_values_at_T_2[2:length_of_series]
V_22 <- X_values_at_T_2[1:(length_of_series - 1)]

# Calculate the correlation between V_1 and V_2
correlation_V1_V2_1 <- cor(V_11, V_12)
correlation_V1_V2_2 <- cor(V_21, V_22)

# Display the correlation
print(paste("Correlation between V_1 and V_2 realization 1:", correlation_V1_V2_1))

## [1] "Correlation between V_1 and V_2 realization 1: -0.0337713177447808"

print(paste("Correlation between V_1 and V_2 realization 2:", correlation_V1_V2_2))

## [1] "Correlation between V_1 and V_2 realization 2: -0.0107623813319924"
```