Econometrics III
Assignment Part III

Thao Le (523716)
David Gyaraki (582340)

Contents

```r
# load packages
if(!require(pacman)){install.packages("pacman")}

p_load(devtools,tidyverse,dplyr,ggplot2,latex2exp, broom, foreign,
 ↪  lubridate, stats)

dfAssign_p3 <- as.data.frame(read_csv("data/data_assign_p3.csv"))
```

1 Question 1

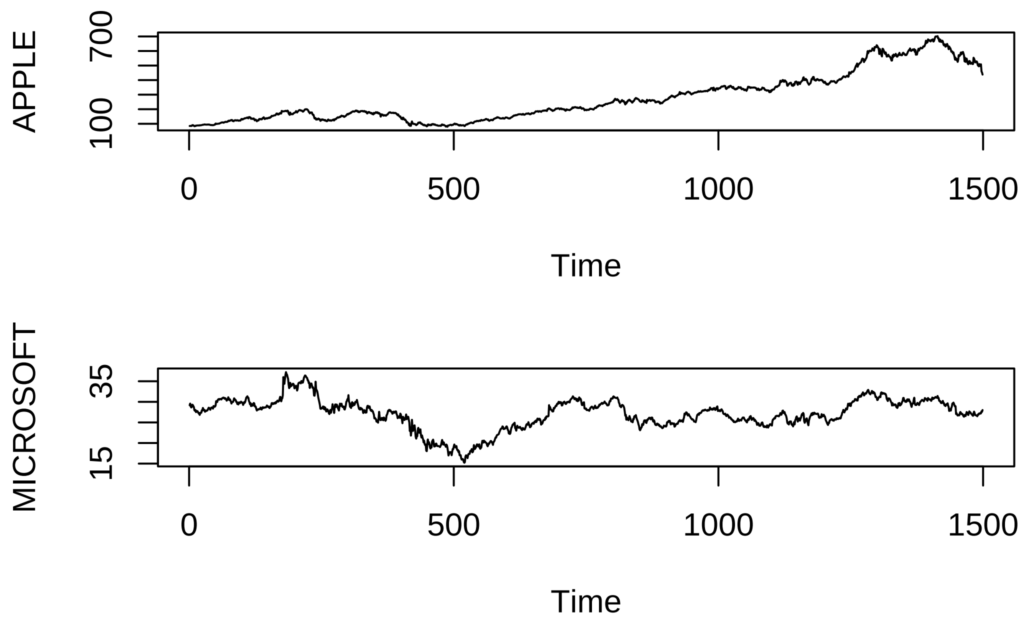
```r
# Set parameter for each walk
sigma_u = 1
sigma_v = 0.5

# Function to simulate X_t, Y_t
simulate_2RW = function(n_steps, sigma_u, sigma_v){
  # Set the starting position for each walk
  start_pos1 <- 0
  start_pos2 <- 0

  # Simulate the random steps for each walk and store in vectors
  steps1 <- rnorm(n_steps, mean = 0, sd = sigma_u)
  steps2 <- rnorm(n_steps, mean = 0, sd = sigma_v)
  Y <- cumsum(steps1) + start_pos1
  X <- cumsum(steps2) + start_pos2

  return(list(X_t=X,Y_t=Y))
}

# Set simulation parameters
num_sims <- 500
max_T <- 1000
step <- 200
init_step = 200

# Initialize dataframe to store results
sim_results <- data.frame(
  T = numeric(),
```

```r
  beta_hat = numeric(),
  t_stat = numeric(),
  R_squared = numeric(),
  stringsAsFactors = FALSE
)

# For each sample size T from T = 200,400,..., 1000, do 500 simulations
↪  each
for (T in seq(init_step, max_T, step)) {
  lRWs = simulate_2RW(T, sigma_u, sigma_v)
  X_t = lRWs$X_t
  Y_t = lRWs$Y_t

  # Perform linear regression and get summary stats
  model <- lm(Y_t ~ X_t)
  summary <- tidy(model)
  beta_hat <- summary$estimate[2]
  t_stat <- summary$statistic[2]
  R_squared <- summary(model)$r.squared
  # Store results
  sim_results <- sim_results %>%
    add_row(T = T,
            beta_hat = beta_hat,
            t_stat = t_stat,
            R_squared = R_squared)
}
```

```r
# Separate results by T
dfT_200 = sim_results[sim_results$T == 200,]
dfT_400 = sim_results[sim_results$T == 400,]
dfT_600 = sim_results[sim_results$T == 600,]
dfT_800 = sim_results[sim_results$T == 800,]
dfT_1000 = sim_results[sim_results$T == 1000,]

#I'm not sure how we show beta_hat converges in distribution to a random
↪  variable here. Do we shot that with different T, beta_hat is
↪  different in expectation/variance? Do we even need to separate by T
↪  even?
```

2 Question 2

```r
# Plot both time series
par(mfrow = c(2, 1))
par(mar = c(4, 4, 2, 1) + 0.1)  # Adjust margins
plot(dfAssign_p3$APPLE, type = "l", xlab = "Time", ylab = "APPLE")
plot(dfAssign_p3$MICROSOFT, type = "l", xlab = "Time", ylab =
↪  "MICROSOFT")
```



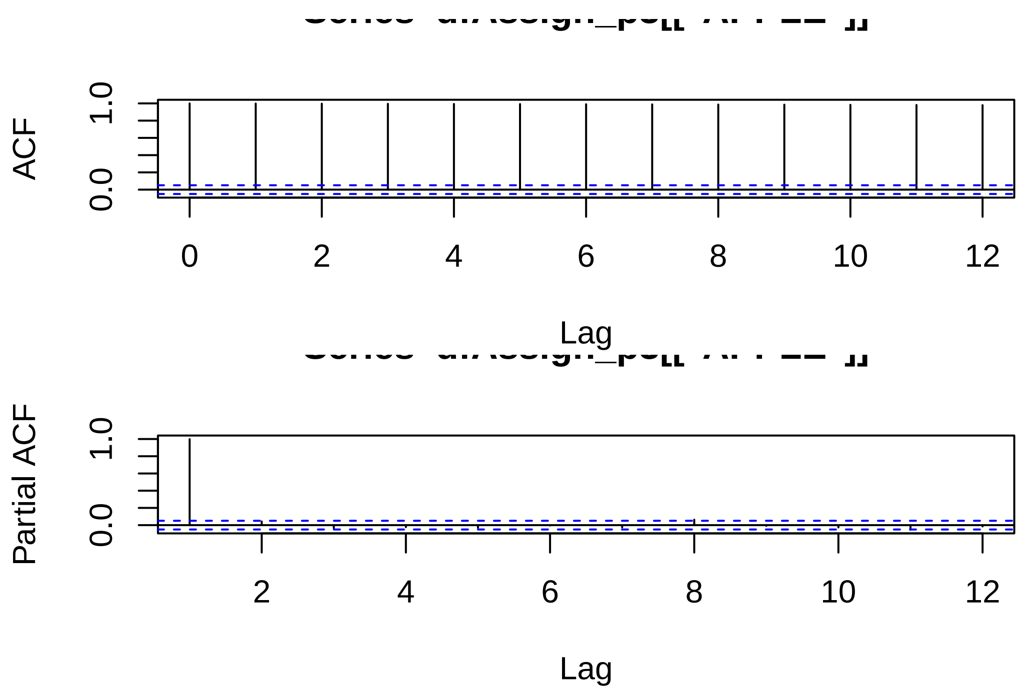

```r
# Calculate and Plot ACF and PACF for Apple
acf(dfAssign_p3[["APPLE"]], 12, pl=F)
```

```
Autocorrelations of series 'dfAssign_p3[["APPLE"]]', by lag

    0     1     2     3     4     5     6     7     8     9    10    11    12 
1.000 0.998 0.997 0.995 0.994 0.992 0.990 0.989 0.987 0.985 0.984 0.982 0.980 
```

```r
acf(dfAssign_p3[["APPLE"]], 12, pl=T)
pacf(dfAssign_p3[["APPLE"]], lag.max = 12, pl=F)
```

```
Partial autocorrelations of series 'dfAssign_p3[["APPLE"]]', by lag

     1      2      3      4      5      6      7      8      9     10     11
 0.998  0.043 -0.037 -0.024 -0.031 -0.005 -0.027  0.064 -0.009 -0.027 -0.054
    12
-0.015
```

```
pacf(dfAssign_p3[["APPLE"]], lag.max = 12, pl=T)
```



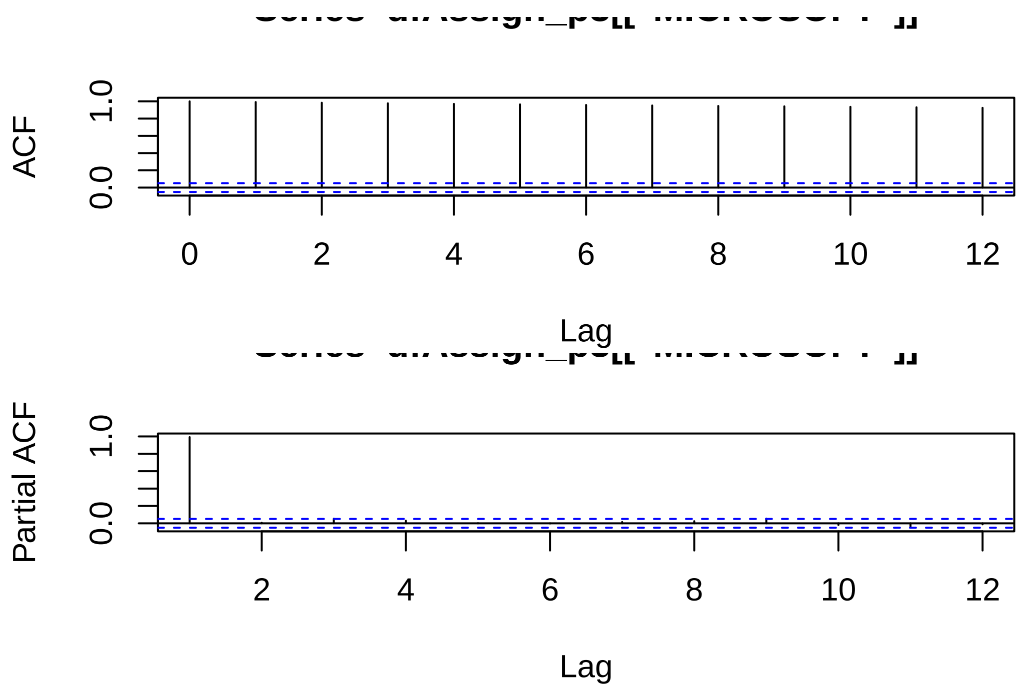

```
# Calculate ACF and PACF for Microsoft
acf(dfAssign_p3[["MICROSOFT"]], 12, pl=F)
```

```
Autocorrelations of series 'dfAssign_p3[["MICROSOFT"]]', by lag

    0     1     2     3     4     5     6     7     8     9    10    11    12
1.000 0.992 0.984 0.978 0.971 0.965 0.959 0.953 0.947 0.942 0.937 0.931 0.925
```

```
acf(dfAssign_p3[["MICROSOFT"]], 12, pl=T)
pacf(dfAssign_p3[["MICROSOFT"]], lag.max = 12, pl=F)
```

```
Partial autocorrelations of series 'dfAssign_p3[["MICROSOFT"]]', by lag

     1      2      3      4      5      6      7      8      9     10     11
 0.992  0.006  0.053  0.031  0.002 -0.002  0.016  0.025  0.054 -0.019 -0.039
    12
-0.012
```

```
pacf(dfAssign_p3[["MICROSOFT"]], lag.max = 12, pl=T)
```



We can see that all 12 lags of the autocorrelation function (ACF) are significant but only one lag of the partial autocorrelation function (PACF) is significant, it suggests that the dynamic of these stocks may have a high degree of autocorrelation but can be adequately modeled using a simple autoregressive (AR) model with one lag.

The high degree of autocorrelation indicated by the significant ACF lags suggests that past values of the stock prices are highly correlated with its current values. This can indicate that the stock price is predictable to some extent and that past values may provide useful information for forecasting future values.

The low number of significant lags in the PACF suggests that the significant ACF lags can be adequately explained by a simple AR(1) model. This implies that the current value of the time series can be explained by its previous value and that other factors such as trend, seasonality, or exogenous variables may not be significant.

We suspect that this also means that the time series can be integrated with order 1.

DO YOU AGREE WITH THIS ANSWER? I'M NOT VERY SURE.

3 Question 3

4 Question 4

5 Question 5