## [ library ]

library("tidyverse")

library("data.table")

library("gridExtra")

library("caret")

library("MLmetrics")

library("lubridate")

library("factoextra")

library("corrplot")

library("Hmisc")

## [ data loading ]

setwd("C:/Users/10sop/Desktop/통계분석학회PSAT/주제분석/데이터/주제분석\_데이터")

raw\_BTC\_price <- "BTC\_KRW coinmarketcap.csv" %>% fread

raw\_exchange\_rate <- "USD\_KRW\_exchange\_rate.csv" %>% fread

raw\_gold\_price <- "Gold\_price.csv" %>% fread

raw\_USA\_bond <- "USA\_10year\_bond\_.csv" %>% fread

raw\_kospi <- "KOSPI.csv" %>% fread

raw\_consumer\_price <- "consumer\_price\_index.csv" %>% fread

## [ data 가공 ]

# 1. 날짜 데이터 형식 맞추기

raw\_BTC\_price[, "날짜"] <- raw\_BTC\_price[, "날짜"] %>% as.matrix %>% dmy

raw\_BTC\_price[, "날짜"]

raw\_exchange\_rate[, 'Date'] <- raw\_exchange\_rate[,'Date'] %>% as.matrix %>% dmy

raw\_exchange\_rate[, 'Date']

raw\_gold\_price[, "Date"] <- raw\_gold\_price[, "Date"] %>% as.matrix %>% mdy()

raw\_gold\_price[, "Date"]

raw\_USA\_bond[, "날짜"] <- raw\_USA\_bond[, "날짜"] %>% as.matrix %>% ymd()

raw\_USA\_bond[,'날짜']

raw\_kospi[, "날짜"] <- raw\_kospi[, "날짜"] %>% as.matrix %>% ymd()

raw\_kospi[, "날짜"]

consumer\_price <- raw\_consumer\_price[-1, c(1,2)]

names(consumer\_price) <- c("연-월 날짜", "총 지수")

consumer\_price[, '연-월 날짜'] %>% as.matrix %>% ym()

consumer\_price[, '연-월 날짜'] <- consumer\_price[, '연-월 날짜'] %>% as.matrix %>% ym()

consumer\_price

## [ 필요한 데이터만 추출하기 - scale-free를 위해 가격이 아닌 '변화율' 선택.]

# 종속변수: 비트코인 일일 수익률

a <- gsub("\\\\",'', raw\_BTC\_price[, '시가\*'] %>% as.matrix) # 시가 안의 \\ 삭제.

b <- gsub(",", "", a) # 시가 안의 ',' 삭제.

raw\_BTC\_price[, "시가\*"] <- b

BTC\_price <- raw\_BTC\_price[, c("날짜", "시가\*")]

BTC\_price[, '시가\*'] <- BTC\_price[,'시가\*'] %>% as.matrix %>% as.numeric

store\_vector <- rep(NA, dim(BTC\_price)[1])

for (i in 1:dim(BTC\_price)[1]) {

store\_vector[i] <- ((BTC\_price[i, "시가\*" ] - BTC\_price[i+1, "시가\*"])/BTC\_price[i+1, "시가\*"] )

} # 일일 수익률 계산

BTC\_merged\_data <- BTC\_price %>% mutate('BTC\_일일\_수익률\_Percent' = as.matrix(store\_vector))

BTC\_profit\_rate <- BTC\_merged\_data[, c("날짜", "BTC\_일일\_수익률\_Percent")] # BTC 날짜와 일일 수익률 데이터

BTC\_profit\_rate[,"BTC\_일일\_수익률\_Percent"] <- gsub("%", "",BTC\_profit\_rate[, "BTC\_일일\_수익률\_Percent"] %>% as.matrix)

BTC\_profit\_rate[,"BTC\_일일\_수익률\_Percent"] <- BTC\_profit\_rate[, "BTC\_일일\_수익률\_Percent"] %>% as.matrix %>% as.numeric

BTC\_profit\_rate

# 변수 (1) USD 환율

exchange\_rate <- raw\_exchange\_rate[, c("Date", "Change %")]

names(exchange\_rate) <- c("날짜", "환율\_변화율\_Percent")

exchange\_rate[,"환율\_변화율\_Percent"] <- gsub("%", "", exchange\_rate[, "환율\_변화율\_Percent"] %>% as.matrix)

exchange\_rate[,"환율\_변화율\_Percent"] <- exchange\_rate[, "환율\_변화율\_Percent"] %>% as.matrix %>% as.numeric

exchange\_rate

# 변수 (2) gold 가격 변화율

store\_vector <- rep(NA, dim(raw\_gold\_price)[1])

for (i in 1:dim(raw\_gold\_price)[1]) {

store\_vector[i] <- ((raw\_gold\_price[i,"Close/Last" ] - raw\_gold\_price[i+1, "Close/Last"])/raw\_gold\_price[i+1, "Close/Last"] )

} # gold 가격 변화율 계산

raw\_gold\_data <- raw\_gold\_price %>% mutate(change\_rate = as.matrix(100\*(store\_vector %>% unlist))) # gold 가격 변화율을 기존의 데이터프레임에 추가하기. %이므로 100을 곱함.

gold\_price\_rate <- raw\_gold\_data[, c("Date", "change\_rate")] # 날짜변수와 gold 가격 변화율을 추출하기 --> 변수 (2) gold 가격 변화율

names(gold\_price\_rate) <- c("날짜", "금\_가격\_변화율\_Percent")

gold\_price\_rate

# 변수 (3) 미국 채권 수익률

USA\_bond <- raw\_USA\_bond[, c("날짜", "변동 %")]

names(USA\_bond) <- c("날짜", "미국\_채권\_수익률\_Percent")

USA\_bond[,"미국\_채권\_수익률\_Percent"] <- gsub("%", "", USA\_bond[, "미국\_채권\_수익률\_Percent"] %>% as.matrix)

USA\_bond[,"미국\_채권\_수익률\_Percent"] <- USA\_bond[, "미국\_채권\_수익률\_Percent"] %>% as.matrix %>% as.numeric

USA\_bond

# 변수 (4) KOSPI 수익률

kospi <- raw\_kospi[, c("날짜", "변동 %")]

names(kospi) <- c("날짜", "KOSPI\_수익률\_Percent")

kospi[,"KOSPI\_수익률\_Percent"] <- gsub("%", "", kospi[, "KOSPI\_수익률\_Percent"] %>% as.matrix)

kospi[,"KOSPI\_수익률\_Percent"] <- kospi[, "KOSPI\_수익률\_Percent"] %>% as.matrix %>% as.numeric

kospi

# 변수 (5) 소비자물가지수

names(consumer\_price) <- c("연-월 날짜", "소비자물가지수")

consumer\_price[,"소비자물가지수"] <- consumer\_price[, "소비자물가지수"] %>% as.matrix %>% as.numeric

## [ 모든 변수 데이터 합치기. ]

# 종속변수;BTC\_profit\_rate: (13.04.29 ~ 21.04.13)

# exchange\_rate: (11.01.03 ~ 21.04.01)

# gold\_price\_rate: (11.04.14 ~ 21.04.13)

# USA\_bond: (11.01.03 ~ 21.04.15)

# kospi: (11.01.03 ~ 21.04.14)

# consumer\_price: (10.01.01 ~ 21.02.01) -> 월별 데이터.

temperary\_frame <- inner\_join(exchange\_rate, gold\_price\_rate, by = "날짜") %>%

inner\_join(USA\_bond) %>%

inner\_join(kospi) %>%

inner\_join(BTC\_profit\_rate) # 일별 데이터는 모두 합쳤음.

temperary\_frame2 <- temperary\_frame %>% mutate(temperary\_frame[, "날짜"] %>% format("%Y-%m")) # 일별 데이터에 월별 데이터를 합치기 위해 작업.

names(temperary\_frame2)[7] <- "연-월 날짜"

temperary\_frame2[, "연-월 날짜"] <- temperary\_frame2[, "연-월 날짜"] %>% as.matrix %>% ym

temperary\_frame3 <- inner\_join(temperary\_frame2, consumer\_price, by = "연-월 날짜") # 일별 데이터의 각 월마다 월별 데이터 값 inner\_join.

temperary\_frame4 <- temperary\_frame3[, -"연-월 날짜"]

final\_frame <- temperary\_frame4[-dim(temperary\_frame4)[1], ] # BTC 일일 수익률을 변수로 쓰므로, 마지막 데이터는 일일 수익률을 구할 수 없음.

# 그래서 NA로 남게 되므로, 그 row 삭제.

final\_frame # 최종 data frame

## [ EDA ]

# 1. 모든 설명변수와 종속변수의 그래프를 한 공간에 그리기 - 추세 확인.

USD\_plot <- ggplot(data = final\_frame,

aes(x = 날짜)) +

geom\_line(aes(y = 환율\_변화율\_Percent), color = "lightblue") +

geom\_line(aes(y = BTC\_일일\_수익률\_Percent\*10), color = 'orange') +

scale\_y\_continuous(

name = "USD Percent axis",

sec.axis = sec\_axis(~. /10, name="BTC Percent axis")

) +

theme\_classic() +

labs(title = "BTC 일일 수익률(%) vs USD 환율 변화율(%)",

x = "날짜",

y = "Percent")+

theme(plot.title = element\_text(hjust = 0.5))

Gold\_plot <- ggplot(data = final\_frame,

aes(x = 날짜)) +

geom\_line(aes(y = 금\_가격\_변화율\_Percent), color = "#CC3399") +

geom\_line(aes(y = BTC\_일일\_수익률\_Percent\*10), color = 'orange') +

scale\_y\_continuous(

name = "Gold Percent axis",

sec.axis = sec\_axis(~. /10, name="BTC Percent axis")

) +

theme\_classic() +

labs(title = "BTC 일일 수익률(%) vs 금 가격 변화율(%)",

x = "날짜",

y = "Percent")+

theme(plot.title = element\_text(hjust = 0.5))

Bond\_plot <- ggplot(data = final\_frame,

aes(x = 날짜)) +

geom\_line(aes(y = 미국\_채권\_수익률\_Percent), color = "green") +

geom\_line(aes(y = BTC\_일일\_수익률\_Percent\*50), color = 'orange') +

scale\_y\_continuous(

name = "Bond Percent axis",

sec.axis = sec\_axis(~. /50, name="BTC Percent axis")

) +

theme\_classic() +

labs(title = "BTC 일일 수익률(%) vs 미국 채권 수익률(%)",

x = "날짜",

y = "Percent") +

theme(plot.title = element\_text(hjust = 0.5))

Kospi\_plot <- ggplot(data = final\_frame,

aes(x = 날짜)) +

geom\_line(aes(y = KOSPI\_수익률\_Percent), color = "pink") +

geom\_line(aes(y = BTC\_일일\_수익률\_Percent\*15), color = 'orange') +

scale\_y\_continuous(

name = "KOSPI Percent axis",

sec.axis = sec\_axis(~. /15, name="BTC Percent axis")

) +

theme\_classic() +

labs(title = "BTC 일일 수익률(%) vs KOSPI 수익률(%)",

x = "날짜",

y = "Percent") +

theme(plot.title = element\_text(hjust = 0.5))

Price\_plot <- ggplot(data = final\_frame,

aes(x = 날짜)) +

geom\_line(aes(y = 소비자물가지수-100),

color = "black",

size = 1.3) +

geom\_line(aes(y = BTC\_일일\_수익률\_Percent\*100),

color = 'orange') +

scale\_y\_continuous(

name = "Price index axis",

sec.axis = sec\_axis(~. /100, name="BTC Percent axis")

) +

theme\_classic() +

labs(title = "BTC 일일 수익률(%) vs 소비자물가지수",

x = "날짜",

y = "Percent") +

theme(plot.title = element\_text(hjust = 0.5))

grid.arrange(USD\_plot, Gold\_plot, Bond\_plot, Kospi\_plot, Price\_plot, nrow= 2)

## 2. 설명변수간 상관관계 확인 --> 상관관계가 크다면 다중회귀분석을 사용하기 곤란하다. 만약 작다면 사용해도 무방하므로 사용해보겠다는 flow

# 피어슨 상관분석 --> 설명변수 간 선형상관관계도 작지만, 설명변수와 종속변수 간 선형상관관계도 너무 작음. 이거 한 후에 비선형 상관관계도 고려해보자.

data\_cor <- cor(final\_frame[,-1])

corrplot(data\_cor, method = 'number', type = "upper")

cor.test(data\_cor,

alternative = c("two.sided"),

method = "pearson")

# 스피어만 상관분석 --> 설명변수와 종속변수 간 비선형상관관계도 너무 작음. 설명변수 추가해야하나.

spearman\_cor <- cor(final\_frame[, -1], method = 'spearman')

corrplot(spearman\_cor, method = 'number', type = 'upper')

# 상관계수 검정. 영가설: 상관계수가 없다.

cor\_pairwise\_test <- rcorr(data\_cor[,-c(1)],

type = 'spearman')

cor\_pairwise\_test[[3]]

corrplot(cor\_pairwise\_test[[3]], method = 'number', type = 'upper') # 상관검정. 모든 설명변수들 상관관계가 유의하지 않다?????

## 3. BTC 가격이 급격히 상승하거나 하락했을 때 설명변수의 경제지표와 관련해서 무슨 사건이 있었는지.

# BTC 가격이 크게 변동했던 상위 10개의 날짜 출력.

sup <- final\_frame[, "BTC\_일일\_수익률\_Percent"] %>% as.matrix %>% sort %>% tail(10)

sup

inf <- final\_frame[, "BTC\_일일\_수익률\_Percent"] %>% as.matrix %>% sort %>% head(10)

inf

sup\_index <- which( final\_frame[, "BTC\_일일\_수익률\_Percent"] %>% as.matrix %in% sup)

sup\_index

inf\_index <-which( final\_frame[, "BTC\_일일\_수익률\_Percent"] %>% as.matrix %in% inf)

inf\_index

final\_frame[sup\_index, "날짜"]

final\_frame[inf\_index, "날짜"]