R Notebook

```
library(here)
setwd(here::here())
library(dplyr)
library(ggplot2)
library(tidyr)
library(purrr)
library(xts)
output_file <- "data/cleaned_etfs.csv"
df <- read.csv(output_file)
df$Date <- as.Date(df[, 1])
data_xts <- xts(df[, -1], order.by = df$Date)
```

```
source("src/stock_list.R")
source("src/generate_dataset.R")
begin_date <- as.Date("2010-01-01")
end_date <- as.Date("2024-05-01")
output_file <- "data/cleaned_etfs.csv"

generate_dataset(stock_namelist, begin_date, end_date, output_file)
```

```
source("src/func_partial_ci.R")
# load csv created in the chunk above

# Define the ticker you want to fit
stock_tickers <- colnames(data_xts)

# Crea le combinazioni di coppie da stimare
stock_pairs <- combn(stock_tickers, 2, simplify = FALSE)

# Parametri di rolling
estimation_years <- 3
rolling_step_months <- 6
save_dir <- "results/fit"

# Esegui il backtest rolling
run_partial_ci_backtest(stock_pairs, data_xts, estimation_years, rolling_step_months, save_dir)
```

```
source("src/filtering_func.R")

results_folder <- "results/fit"
save_dir <- "results/pairs"

# Filter parameters
rho_min <- 0.9
rho_max <- 0.98
rsq_min <- 0.9
loglik_max <- 0
```

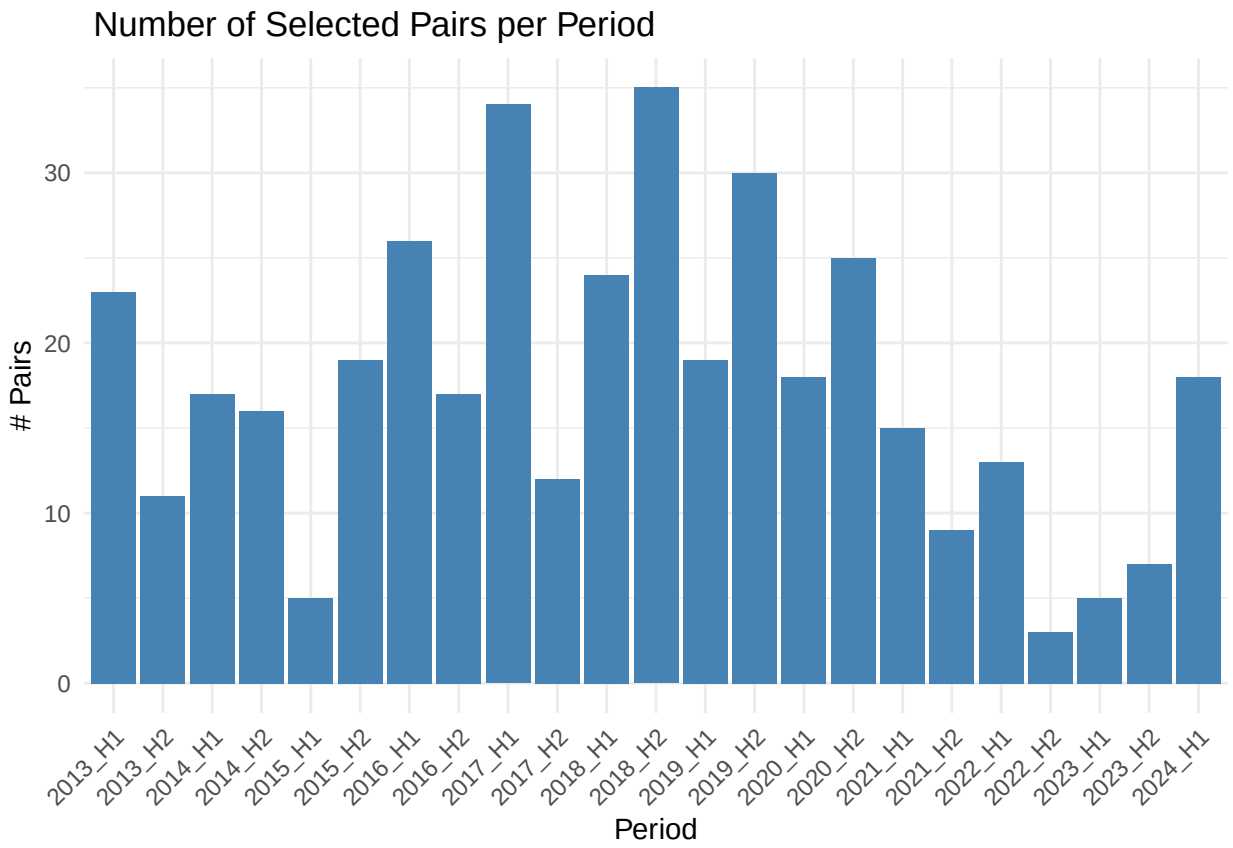
```r
for (year in 2013:2024) {
  for (half in c("H1", "H2")) {
    process_period(year, half,
                   results_folder = results_folder,
                   rho_min = rho_min,
                   rho_max = rho_max,
                   rsq_min = rsq_min,
                   loglik_max = loglik_max,
                   save_dir = save_dir)
  }
}
```

```
##  Saved filtered pairs for 2013_H1 to results/pairs/pairs_2013_H1.RData
##  Saved filtered pairs for 2013_H2 to results/pairs/pairs_2013_H2.RData
##  Saved filtered pairs for 2014_H1 to results/pairs/pairs_2014_H1.RData
##  Saved filtered pairs for 2014_H2 to results/pairs/pairs_2014_H2.RData
##  Saved filtered pairs for 2015_H1 to results/pairs/pairs_2015_H1.RData
##  Saved filtered pairs for 2015_H2 to results/pairs/pairs_2015_H2.RData
##  Saved filtered pairs for 2016_H1 to results/pairs/pairs_2016_H1.RData
##  Saved filtered pairs for 2016_H2 to results/pairs/pairs_2016_H2.RData
##  Saved filtered pairs for 2017_H1 to results/pairs/pairs_2017_H1.RData
##  Saved filtered pairs for 2017_H2 to results/pairs/pairs_2017_H2.RData
##  Saved filtered pairs for 2018_H1 to results/pairs/pairs_2018_H1.RData
##  Saved filtered pairs for 2018_H2 to results/pairs/pairs_2018_H2.RData
##  Saved filtered pairs for 2019_H1 to results/pairs/pairs_2019_H1.RData
##  Saved filtered pairs for 2019_H2 to results/pairs/pairs_2019_H2.RData
##  Saved filtered pairs for 2020_H1 to results/pairs/pairs_2020_H1.RData
##  Saved filtered pairs for 2020_H2 to results/pairs/pairs_2020_H2.RData
##  Saved filtered pairs for 2021_H1 to results/pairs/pairs_2021_H1.RData
##  Saved filtered pairs for 2021_H2 to results/pairs/pairs_2021_H2.RData
##  Saved filtered pairs for 2022_H1 to results/pairs/pairs_2022_H1.RData
##  Saved filtered pairs for 2022_H2 to results/pairs/pairs_2022_H2.RData
##  Saved filtered pairs for 2023_H1 to results/pairs/pairs_2023_H1.RData
##  Saved filtered pairs for 2023_H2 to results/pairs/pairs_2023_H2.RData
##  Saved filtered pairs for 2024_H1 to results/pairs/pairs_2024_H1.RData
##  File not found: results/fit/res_2024_H2.RData
```

```r
pairs_dir <- "results/pairs/"

#  list all the pairs
pair_files <- list.files(pairs_dir, pattern = "^pairs_.*\\.RData$", full.names = TRUE)

all_pairs <- list()

for (file in pair_files) {
  temp_env <- new.env()
  load(file, envir = temp_env)
  var_name <- ls(temp_env)[grepl("^pairs_", ls(temp_env))]
  pairs <- get(var_name, envir = temp_env)
  all_pairs[[gsub("pairs_|\\.RData", "", basename(file))]] <- pairs
}

# Rbind all the pairs in a unique dataset
pairs_df <- do.call(rbind, lapply(names(all_pairs), function(period) {
```

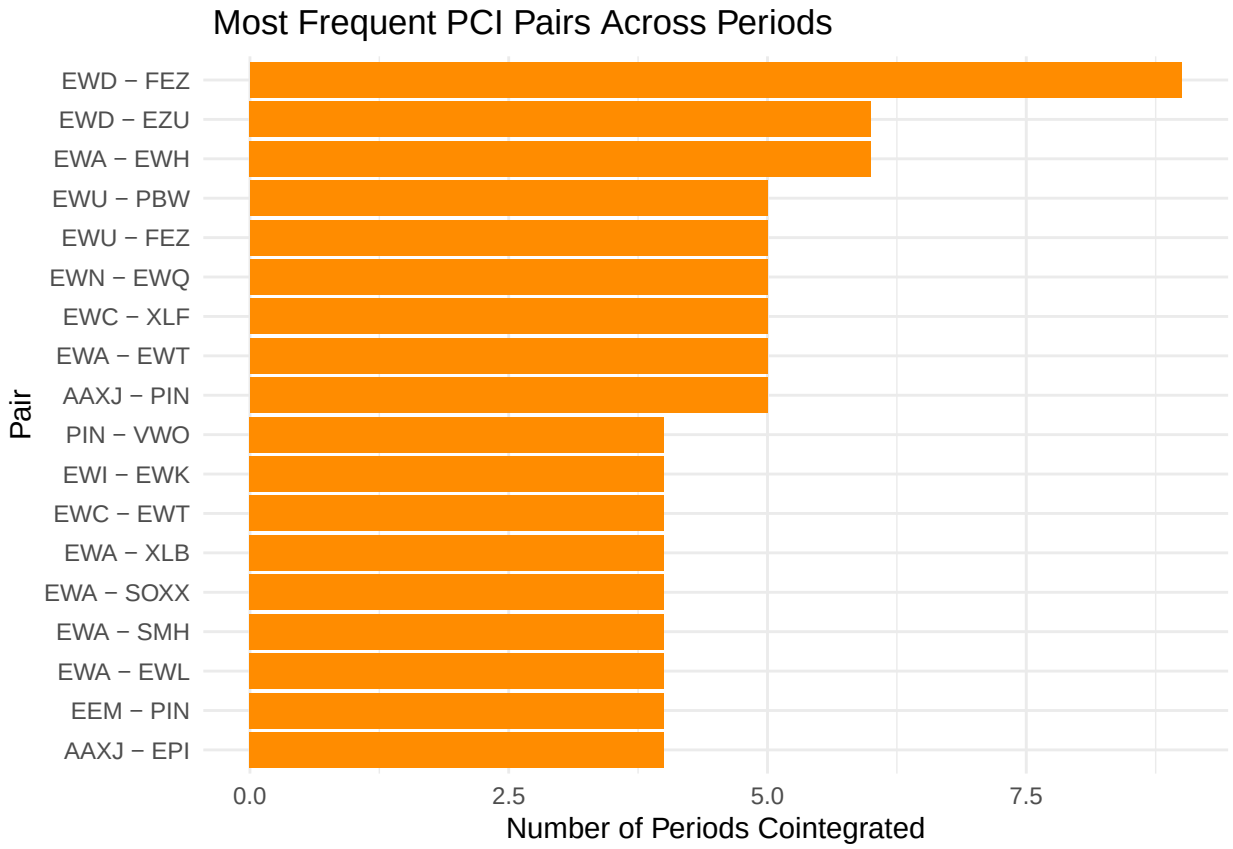
```
  do.call(rbind, lapply(all_pairs[[period]], function(pair) {
    data.frame(period = period, stock_a = pair[1], stock_b = pair[2])
  }))
}))
```

```
# Count the number of selected pairs for each period
pair_counts <- pairs_df %>%
  group_by(period) %>%
  summarise(num_pairs = n())

ggplot(pair_counts, aes(x = period, y = num_pairs)) +
  geom_col(fill = "steelblue") +
  theme_minimal() +
  labs(title = " Number of Selected Pairs per Period",
       x = "Period",
       y = "# Pairs") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

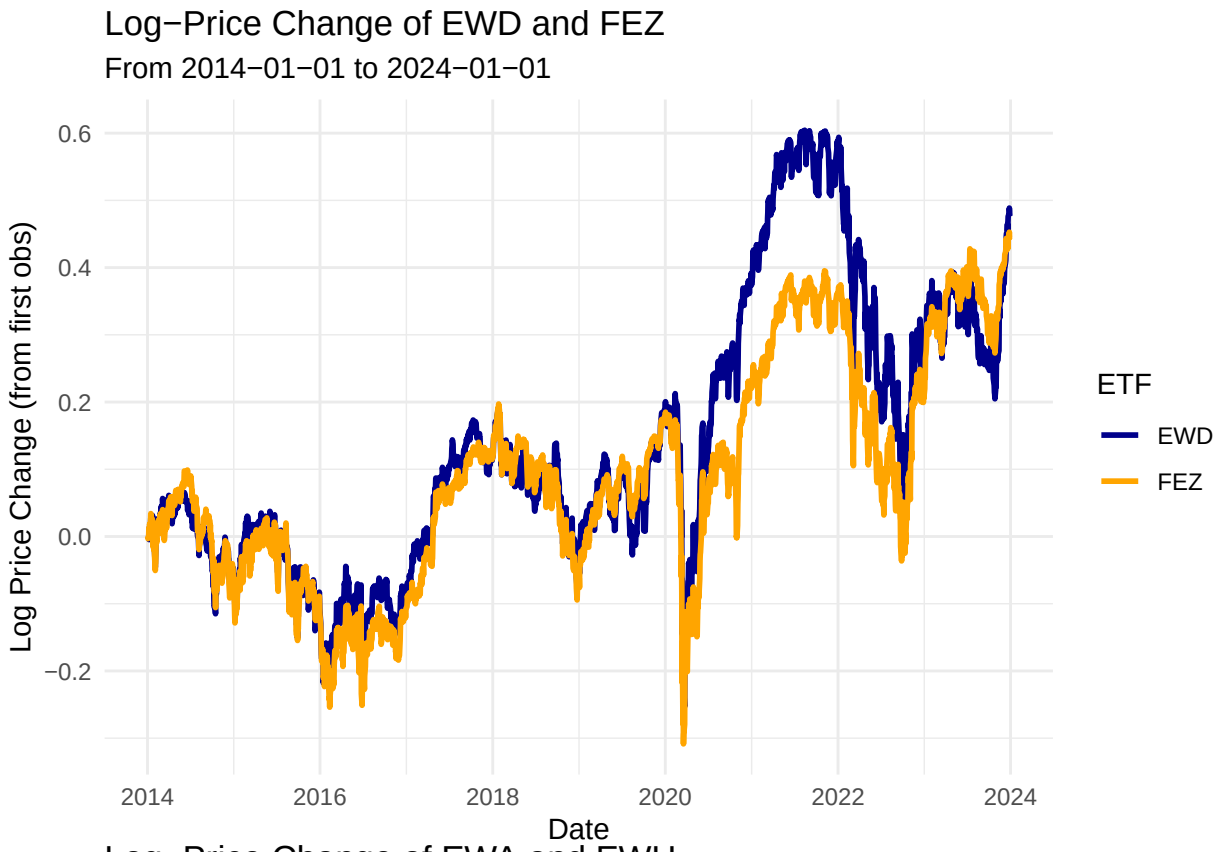


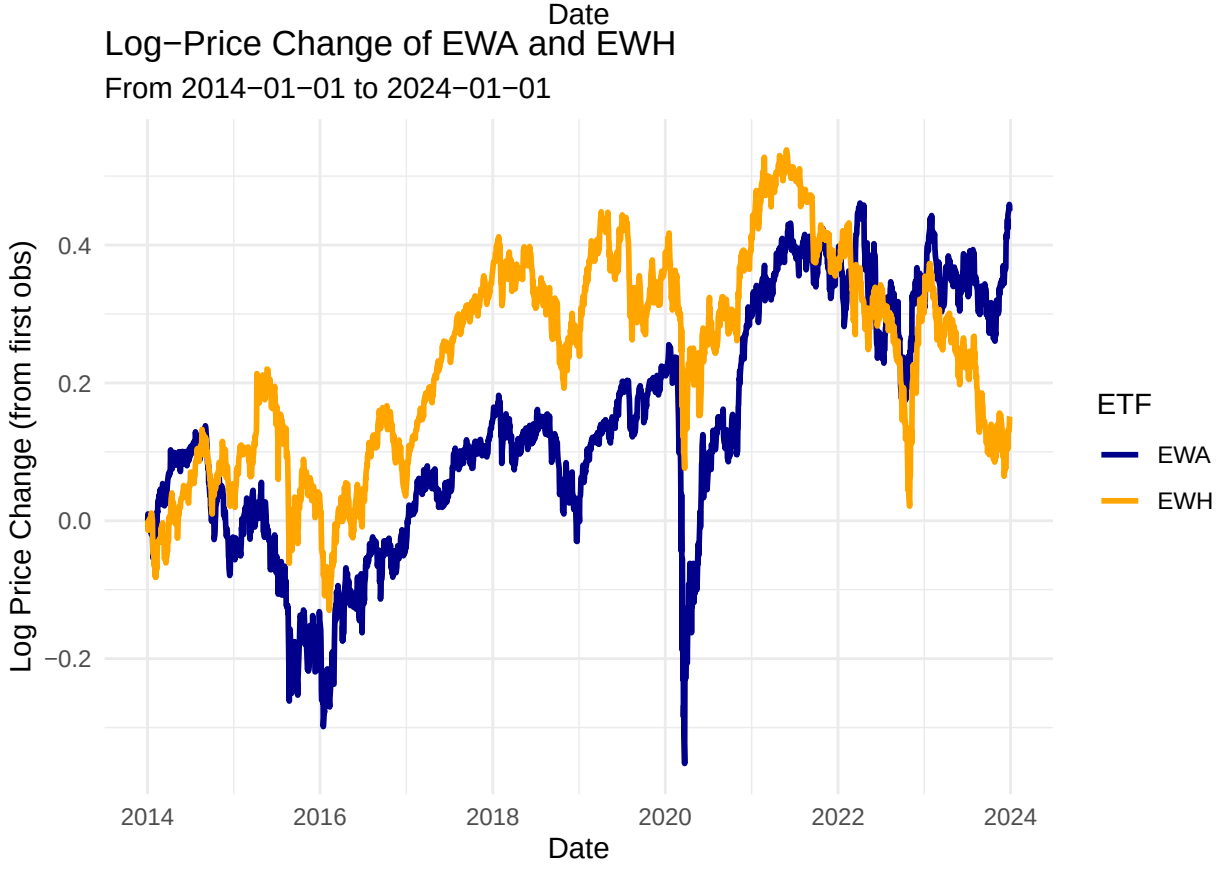

```
#  Crea nome univoco per ogni coppia (ordine alfabetico per evitare duplicati invertiti)
pairs_df <- pairs_df %>%
  mutate(pair = paste(pmin(stock_a, stock_b), pmax(stock_a, stock_b), sep = " - "))

#  Conta le ricorrenze per ciascuna coppia
top_pairs <- pairs_df %>%
  count(pair, sort = TRUE) %>%
  top_n(10, n)
```

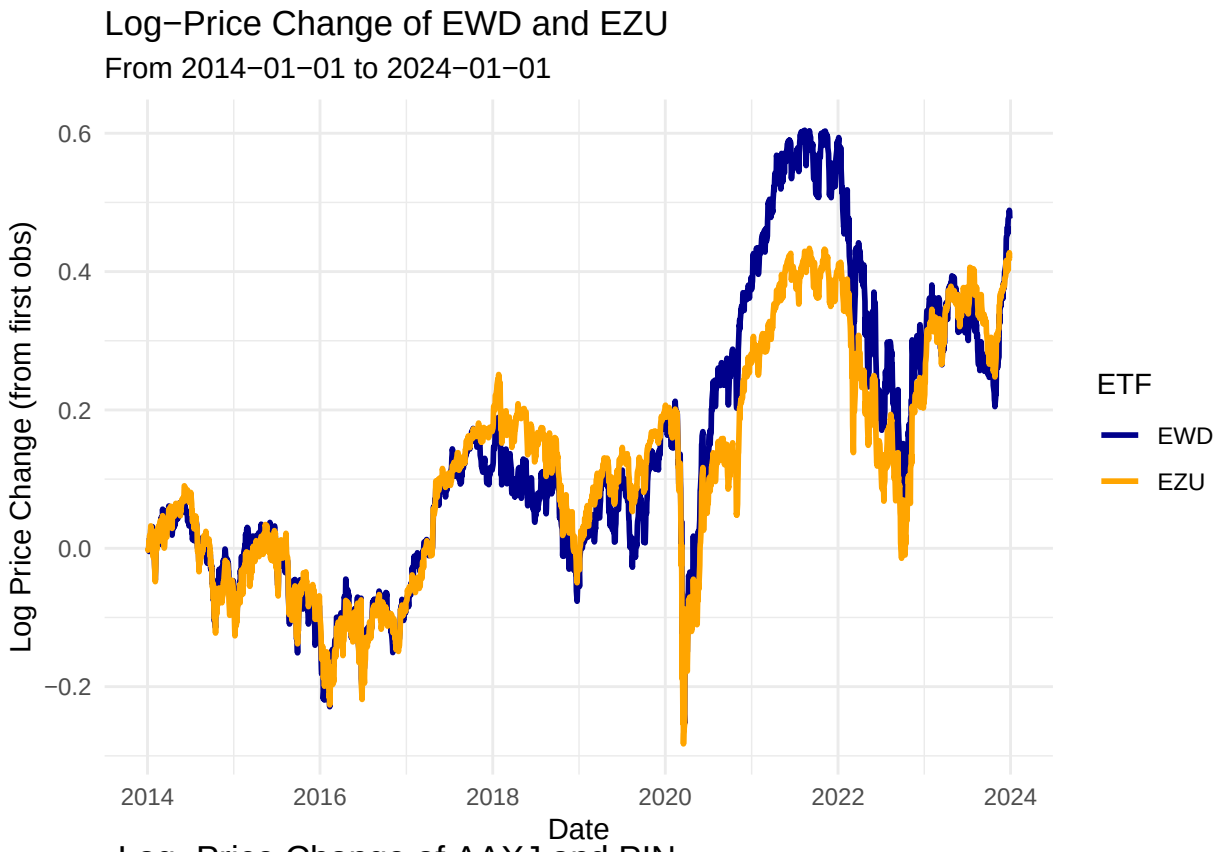
```r
#  Grafico a barre
ggplot(top_pairs, aes(x = reorder(pair, n), y = n)) +
  geom_col(fill = "darkorange") +
  coord_flip() +
  theme_minimal() +
  labs(title = " Most Frequent PCI Pairs Across Periods",
       x = "Pair",
       y = "Number of Periods Cointegrated")
```

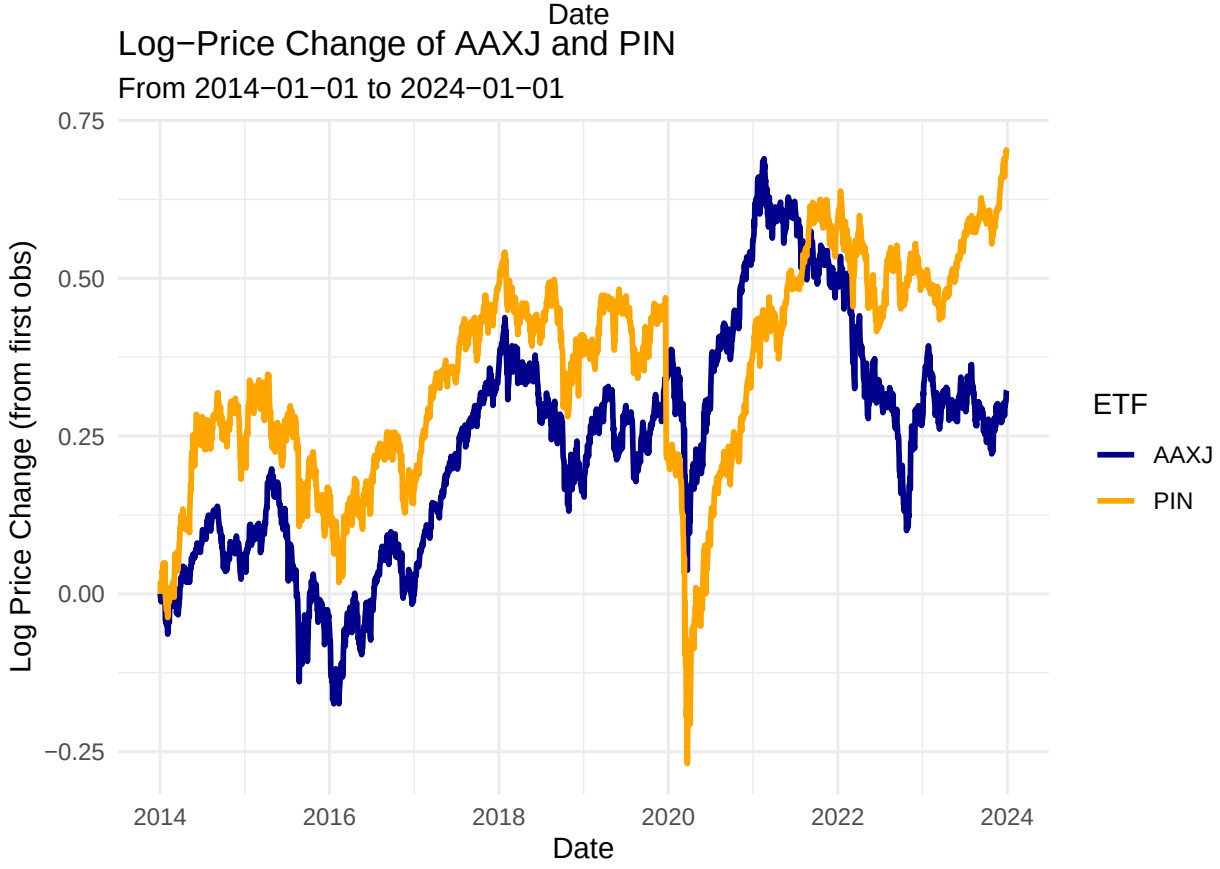


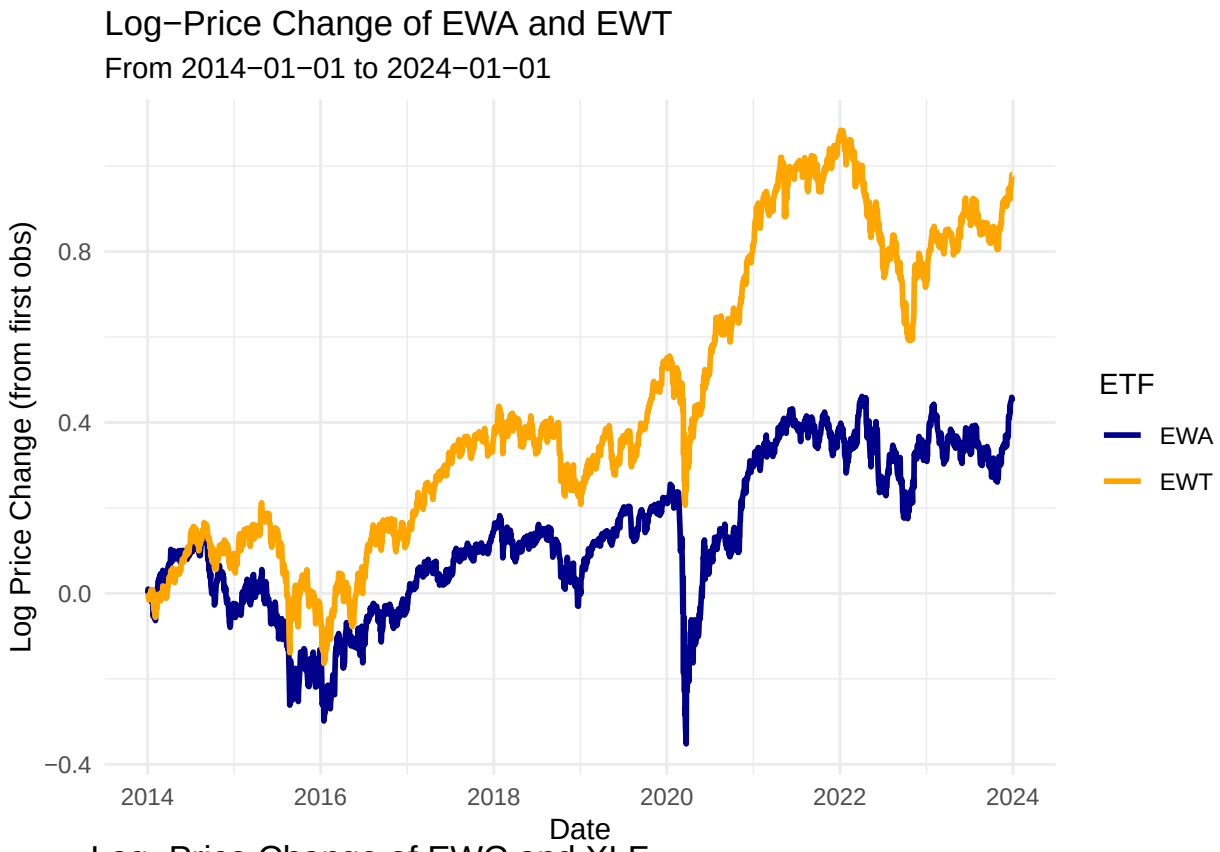

```r
source("src/plots_func.R")

walk(top_pairs$pair, function(p) {
  tickers <- unlist(strsplit(p, " - "))
  print(plot_pair_log_price_change(data_xts, tickers[1], tickers[2],
                                   start_date = "2014-01-01", end_date = "2024-01-01"))
})
```

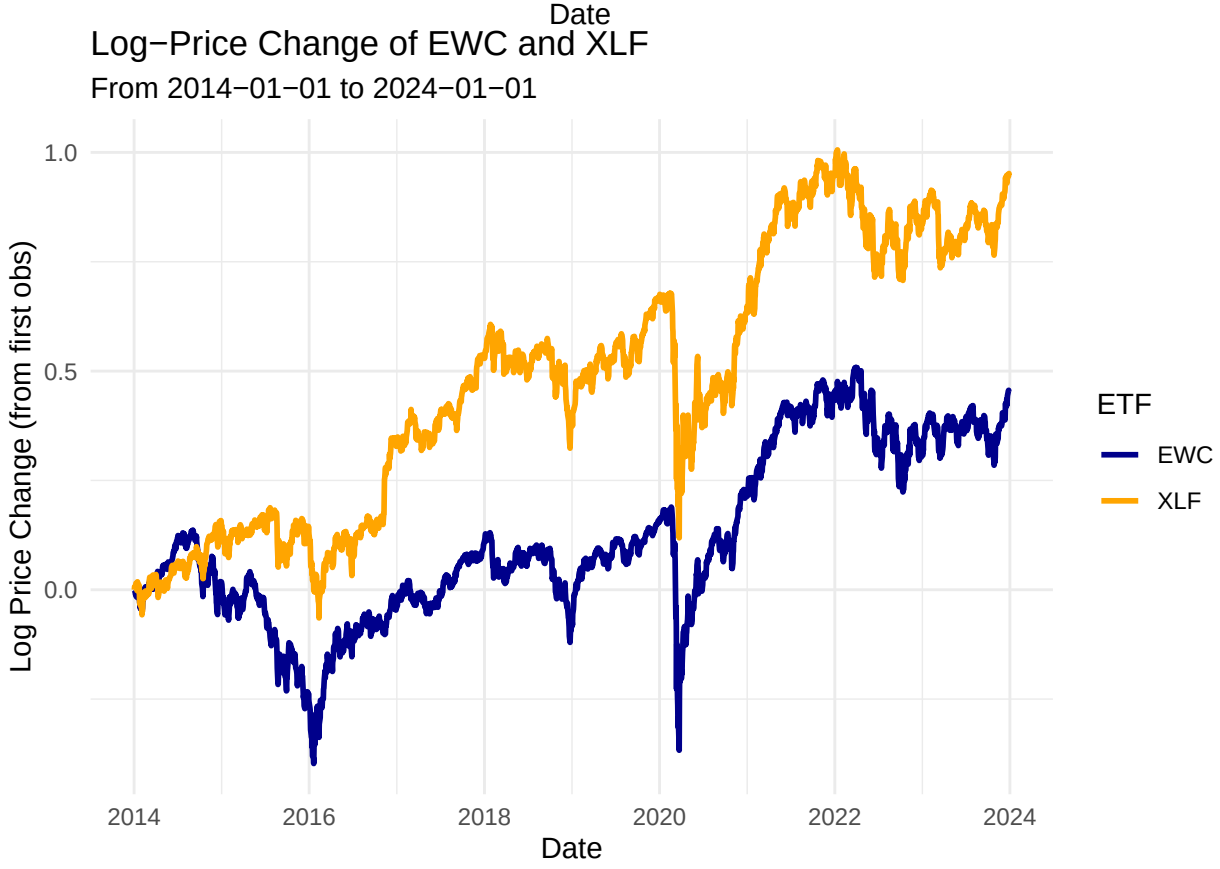

Log−Price Change of EWD and FEZ
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
0.6
0.4
0.2
0.0
−0.2
2014
2016
2018
2020
2022
2024
Date
ETF
EWD
FEZ

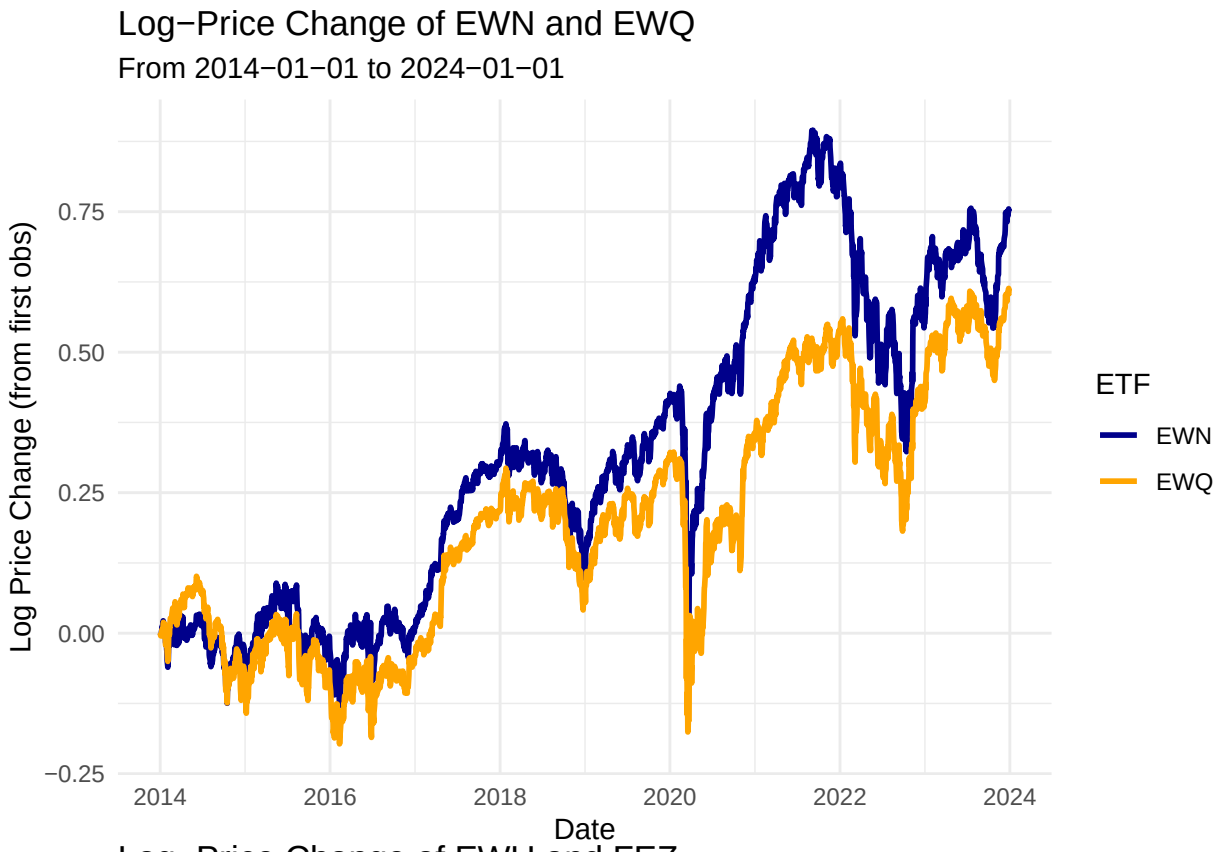

Log−Price Change of EWA and EWH
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
0.4
0.2
0.0
−0.2
2014
2016
2018
2020
2022
2024
Date
ETF
EWA
EWH

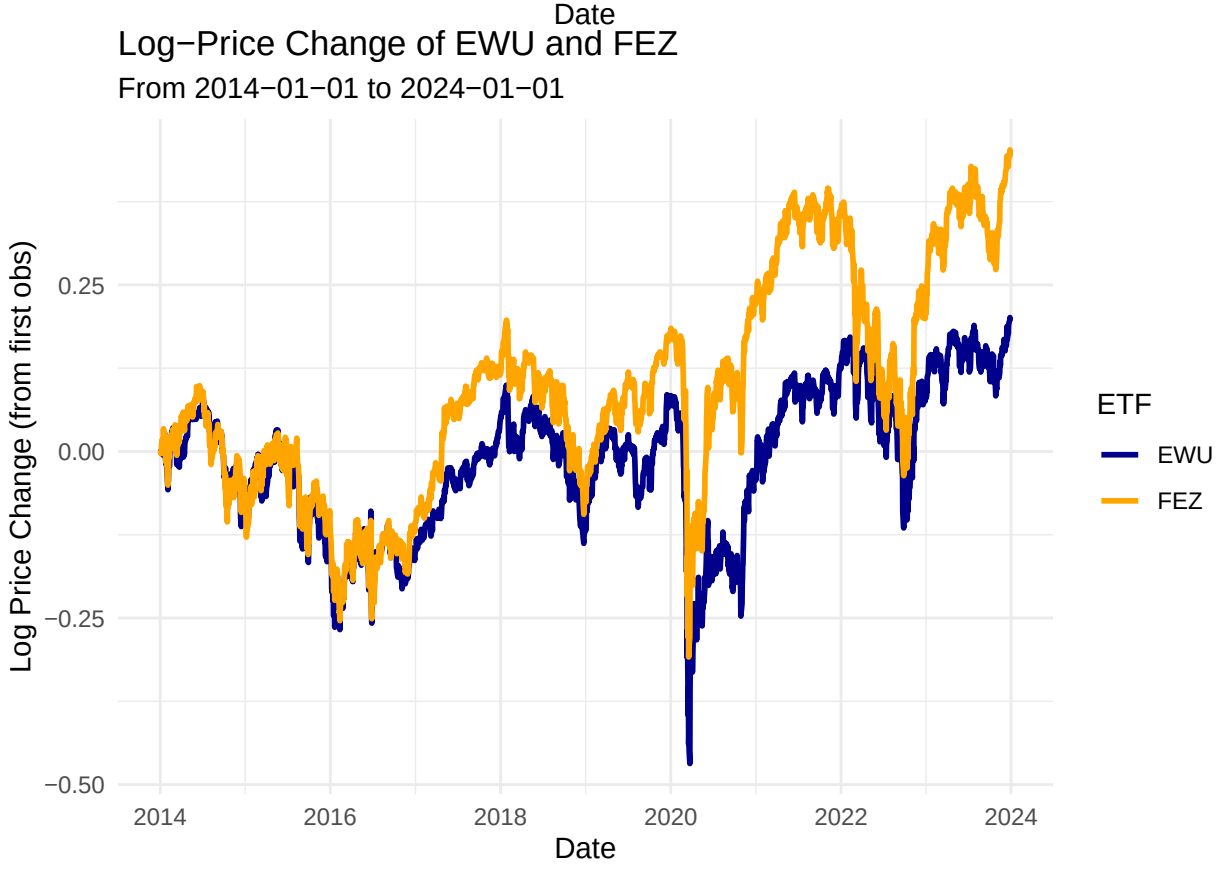

Log−Price Change of EWD and EZU
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
0.6
0.4
0.2
0.0
−0.2
2014
2016
2018
2020
2022
2024
Date
ETF
EWD
EZU

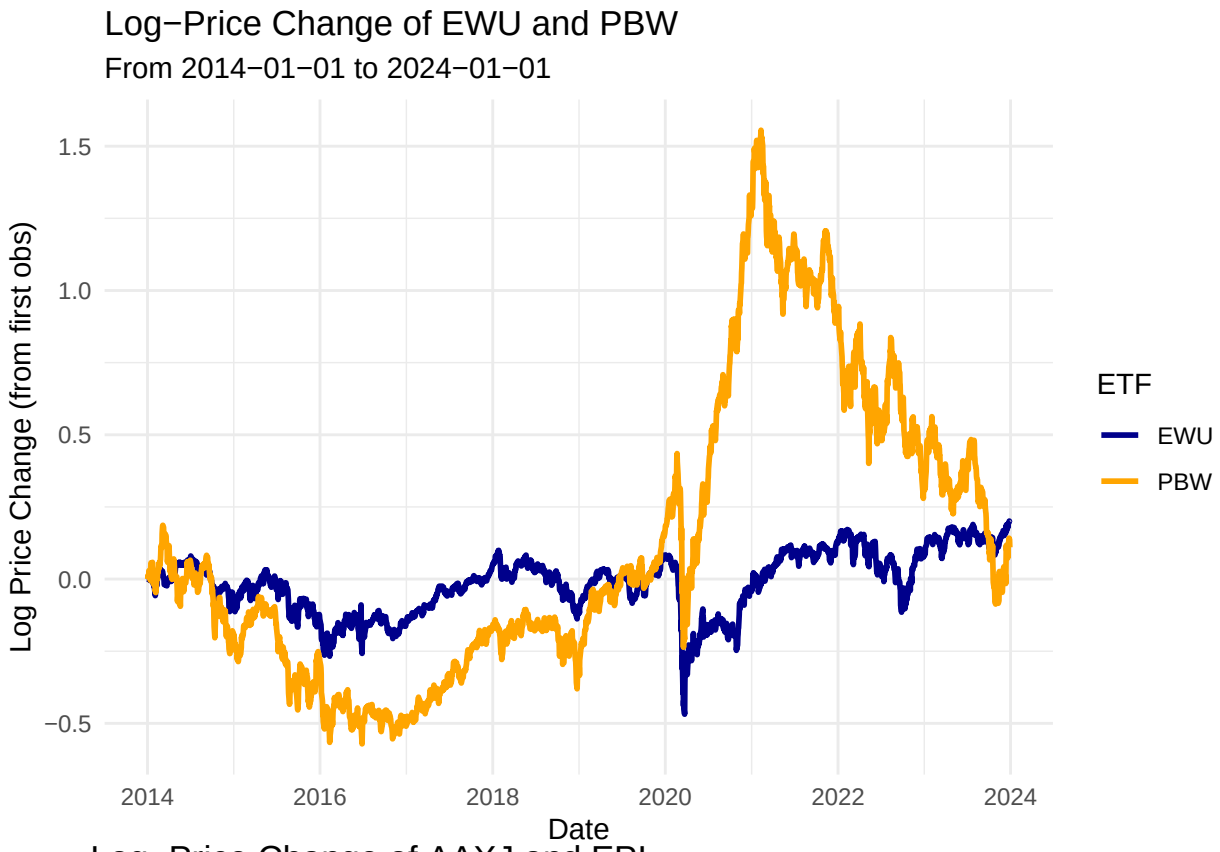

Log−Price Change of AAXJ and PIN
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
0.75
0.50
0.25
0.00
−0.25
2014
2016
2018
2020
2022
2024
Date
ETF
AAXJ
PIN

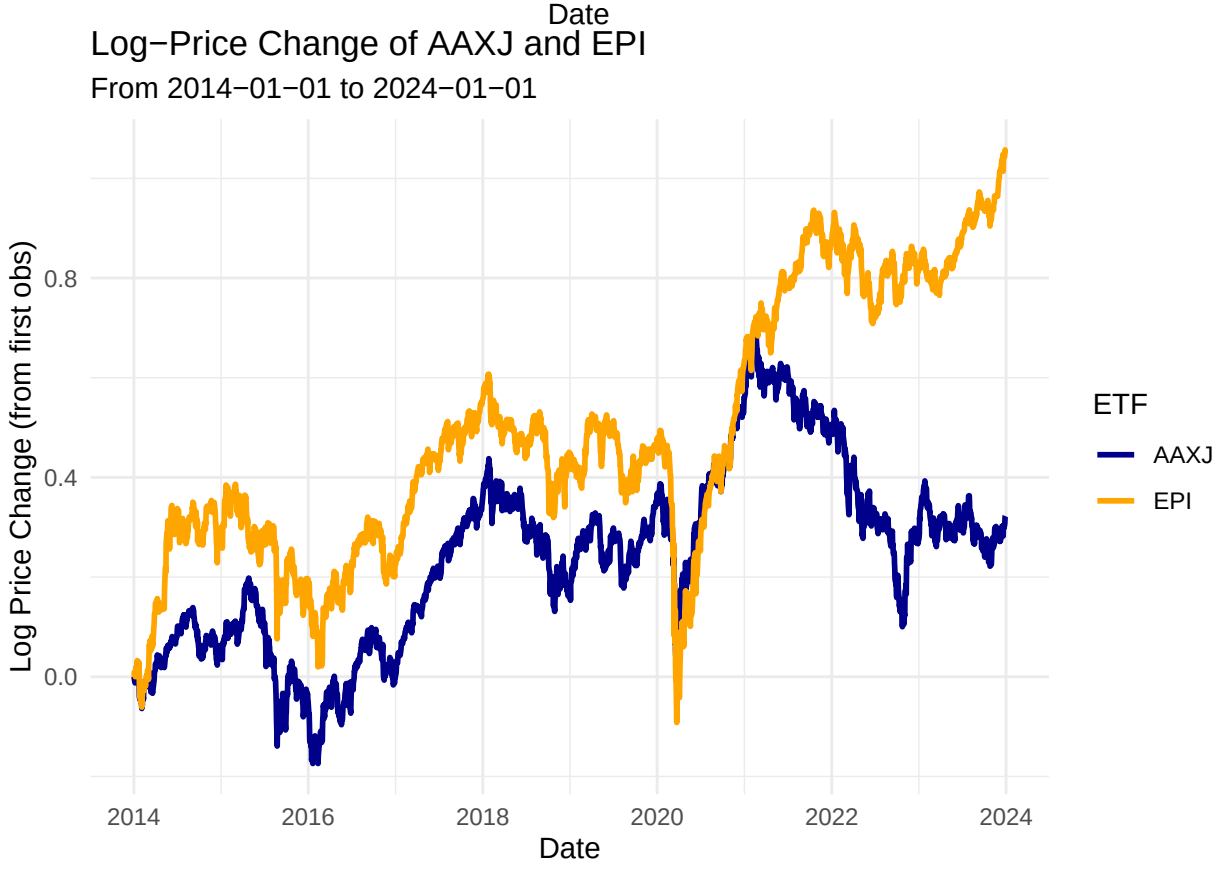

Log−Price Change of EWA and EWT
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
0.8
0.4
0.0
−0.4
2014
2016
2018
2020
2022
2024
Date
ETF
EWA
EWT

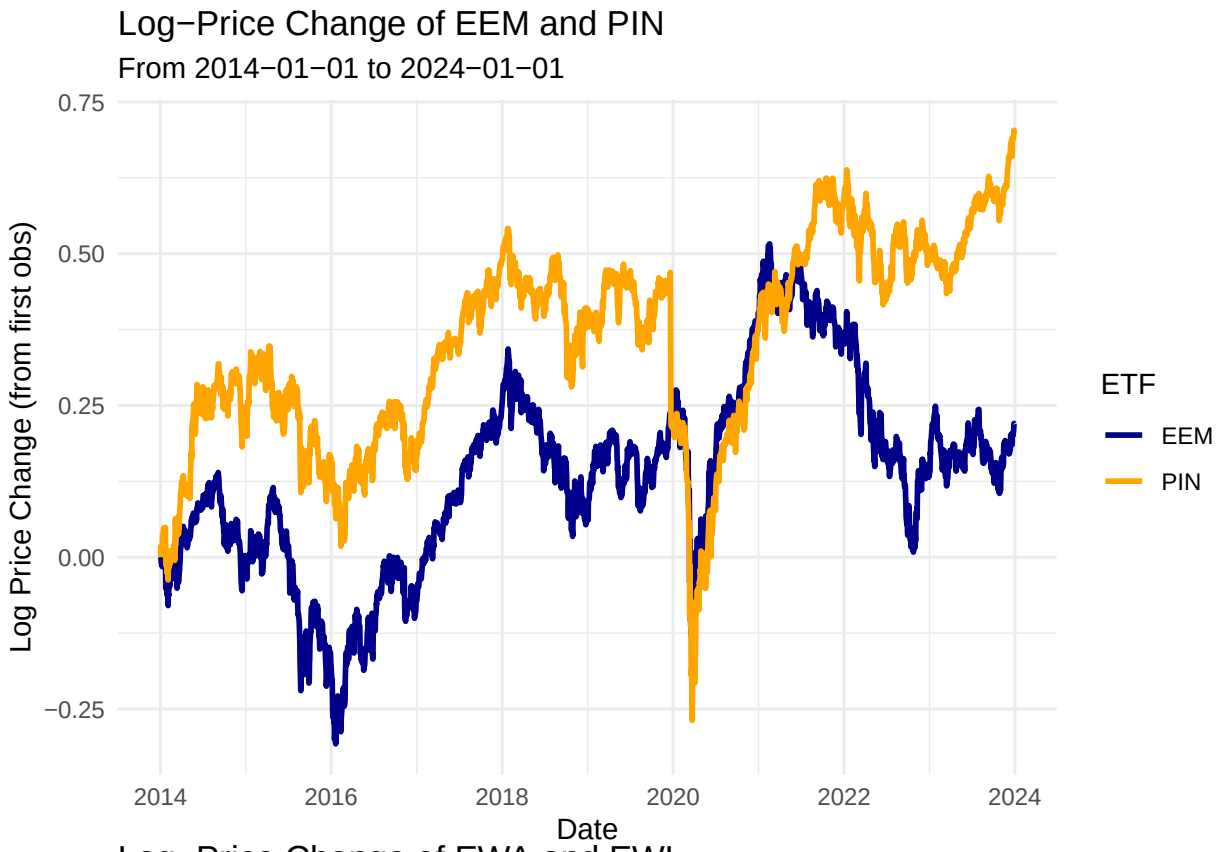

Log−Price Change of EWC and XLF
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
1.0
0.5
0.0
2014
2016
2018
2020
2022
2024
Date
ETF
EWC
XLF

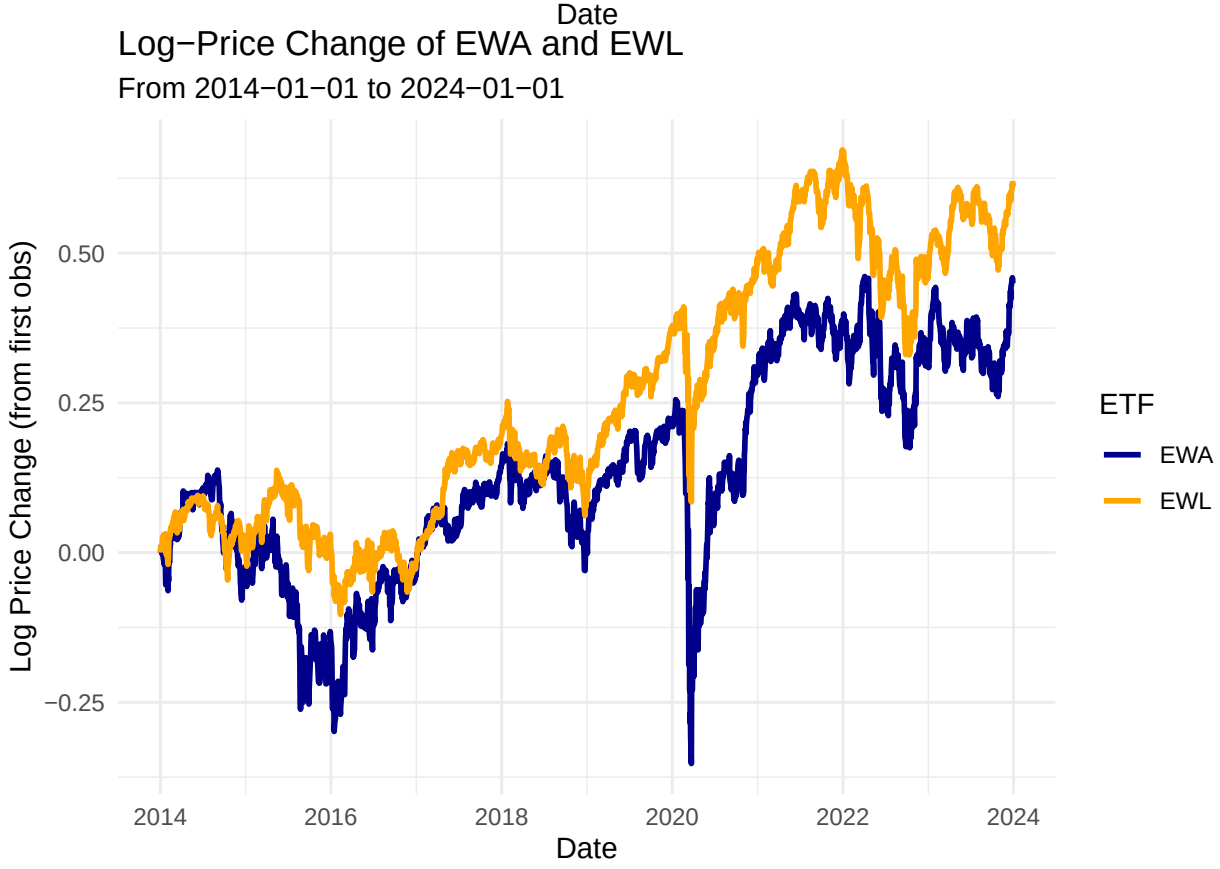

Log−Price Change of EWN and EWQ
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
0.75
0.50
0.25
0.00
−0.25
2014
2016
2018
2020
2022
2024
Date
ETF
EWN
EWQ

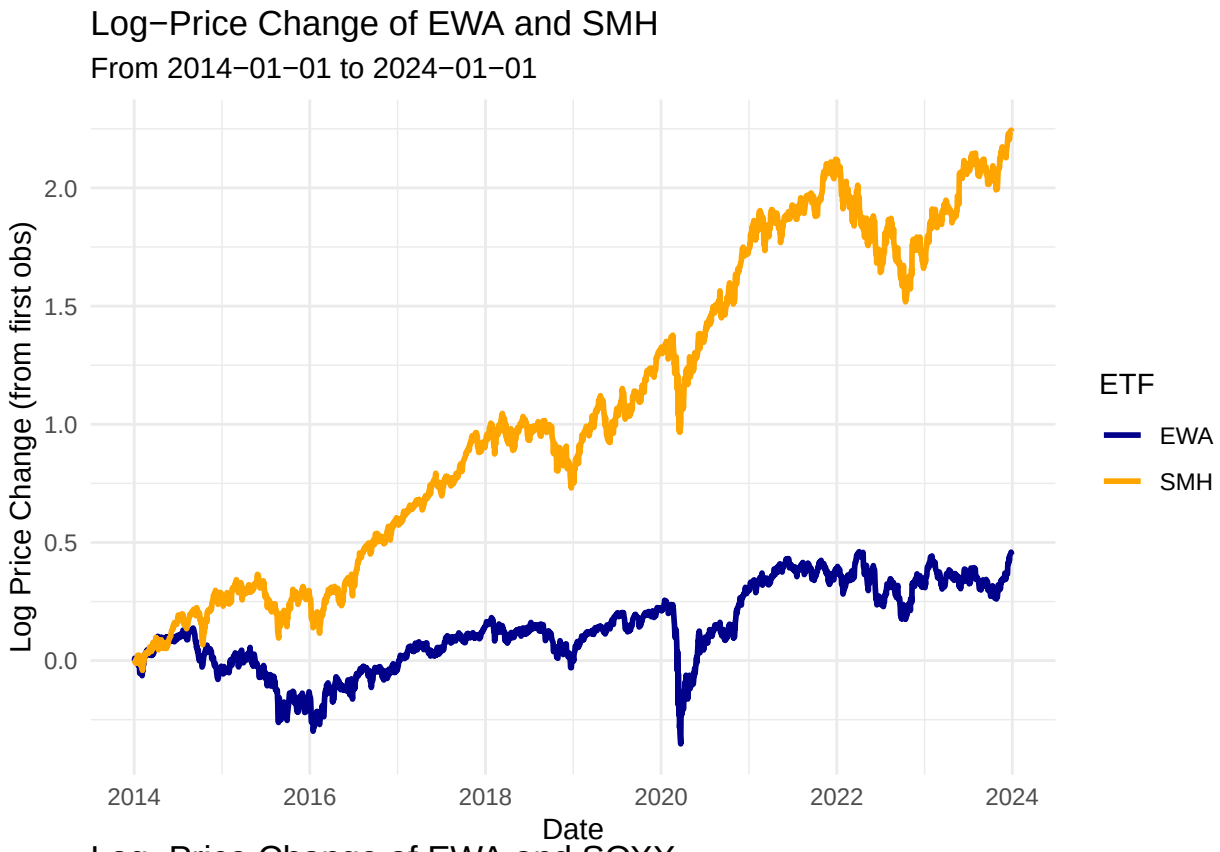

Log−Price Change of EWU and FEZ
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
0.25
0.00
−0.25
−0.50
2014
2016
2018
2020
2022
2024
Date
ETF
EWU
FEZ

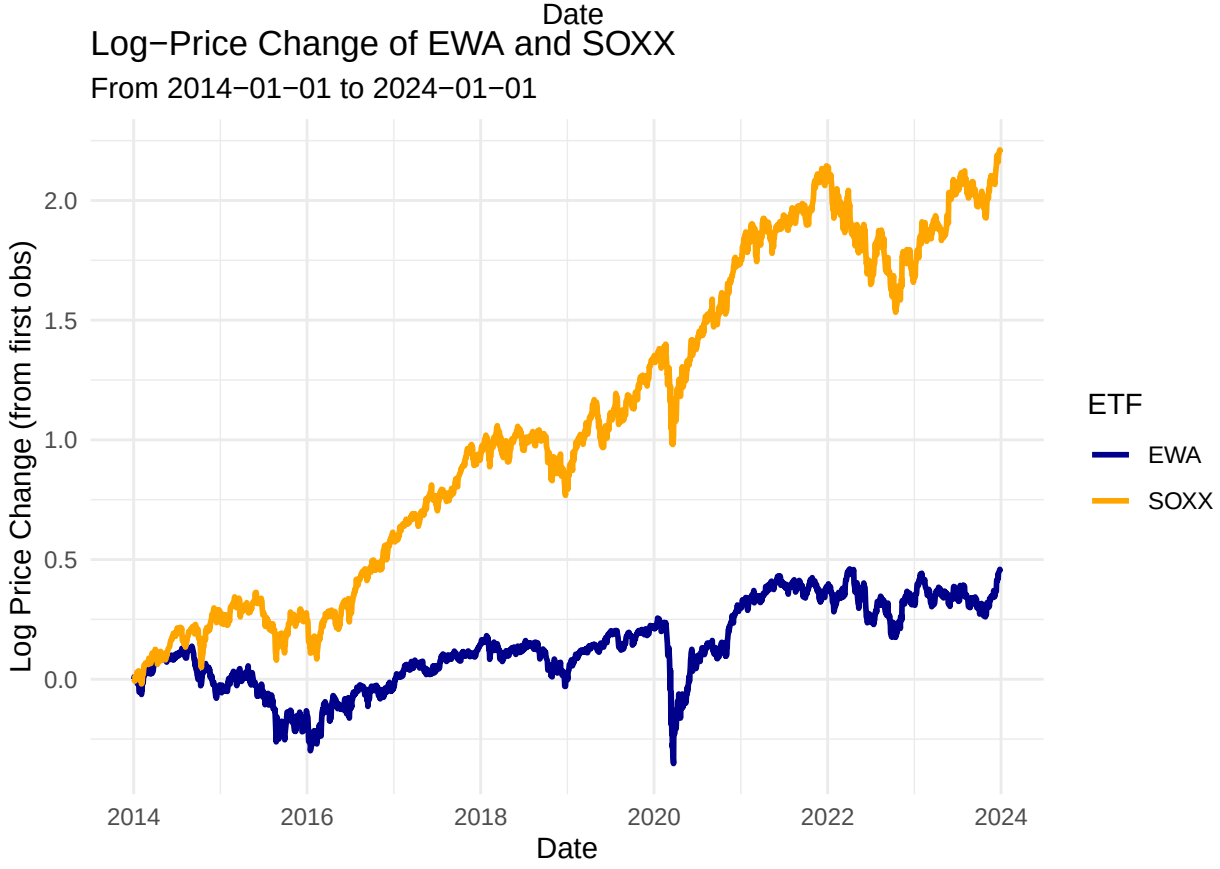

Log−Price Change of EWU and PBW
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
1.5
1.0
0.5
0.0
−0.5
2014
2016
2018
2020
2022
2024
Date
ETF
EWU
PBW

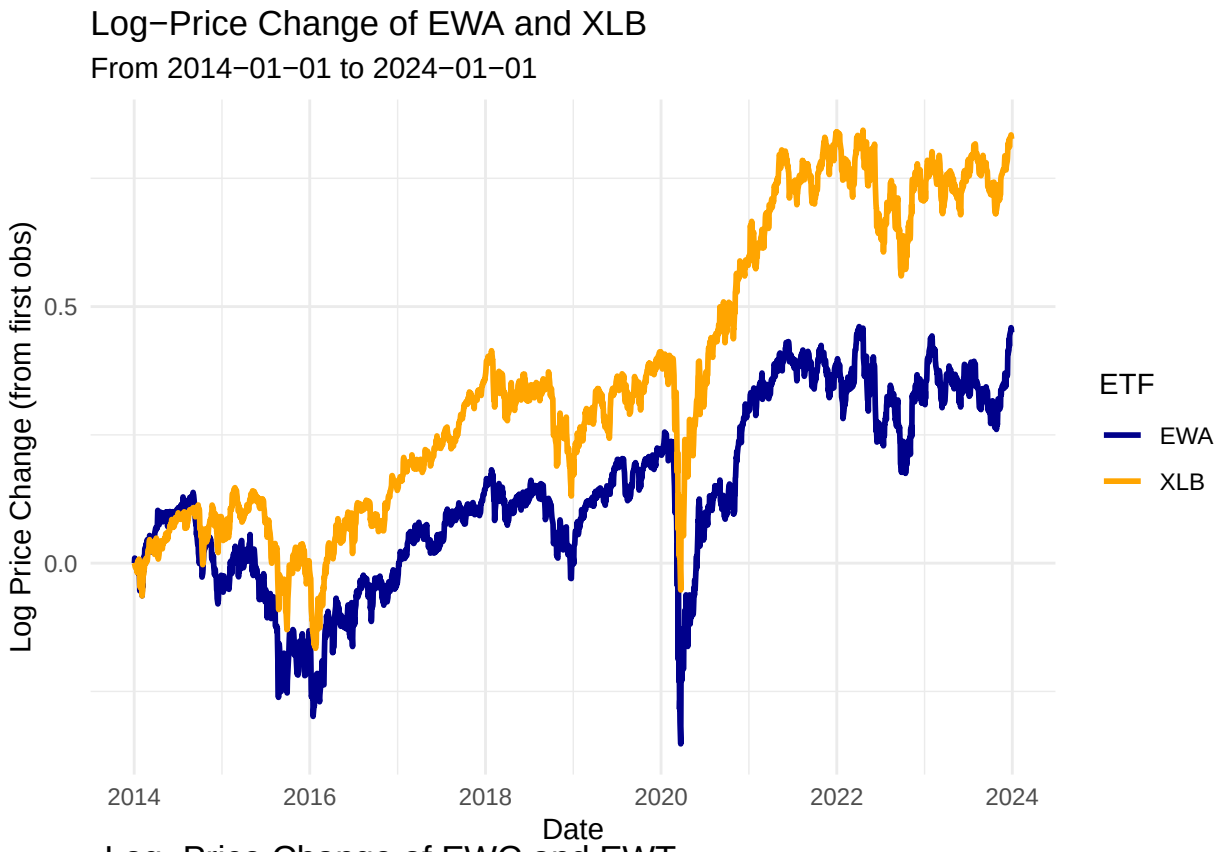

Log−Price Change of AAXJ and EPI
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
0.8
0.4
0.0
2014
2016
2018
2020
2022
2024
Date
ETF
AAXJ
EPI

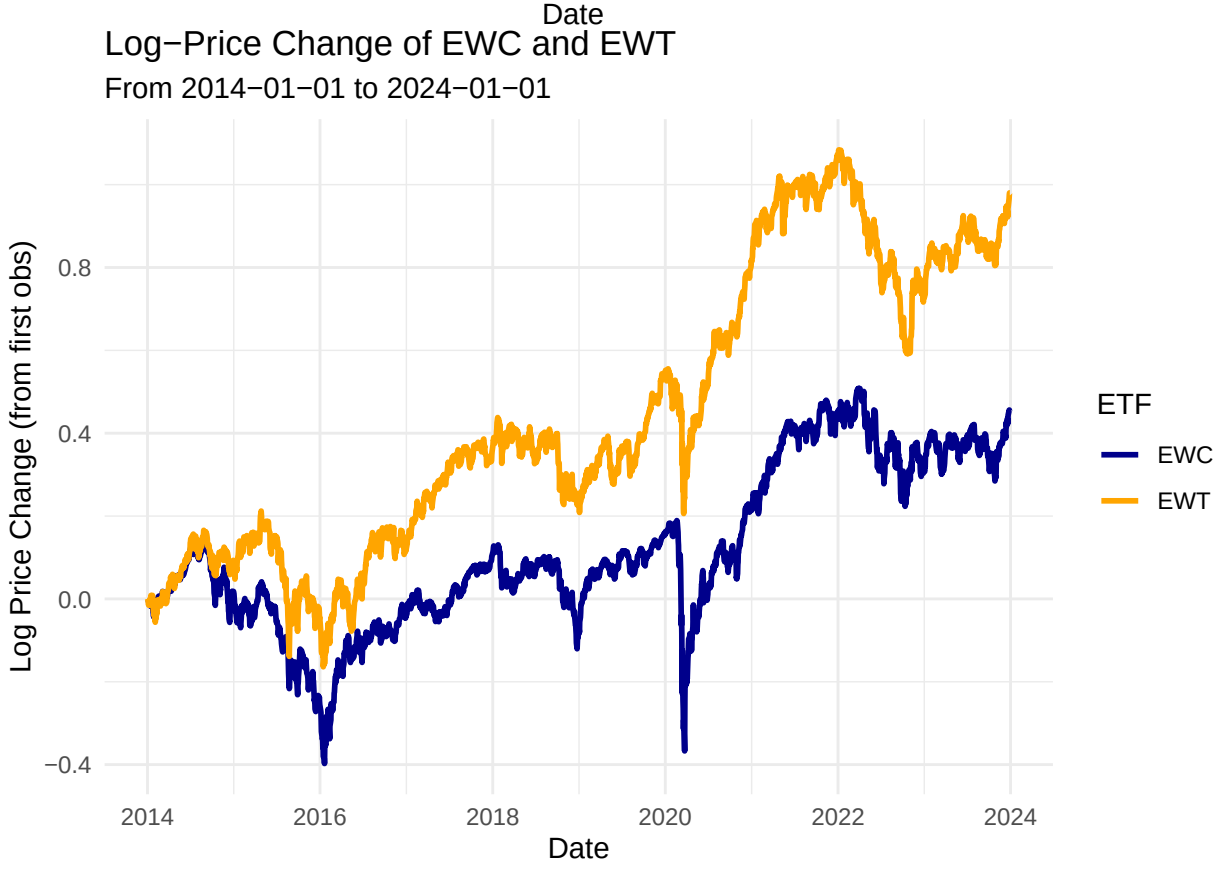

Log−Price Change of EEM and PIN
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
0.75
0.50
0.25
0.00
−0.25
2014
2016
2018
2020
2022
2024
Date
ETF
EEM
PIN

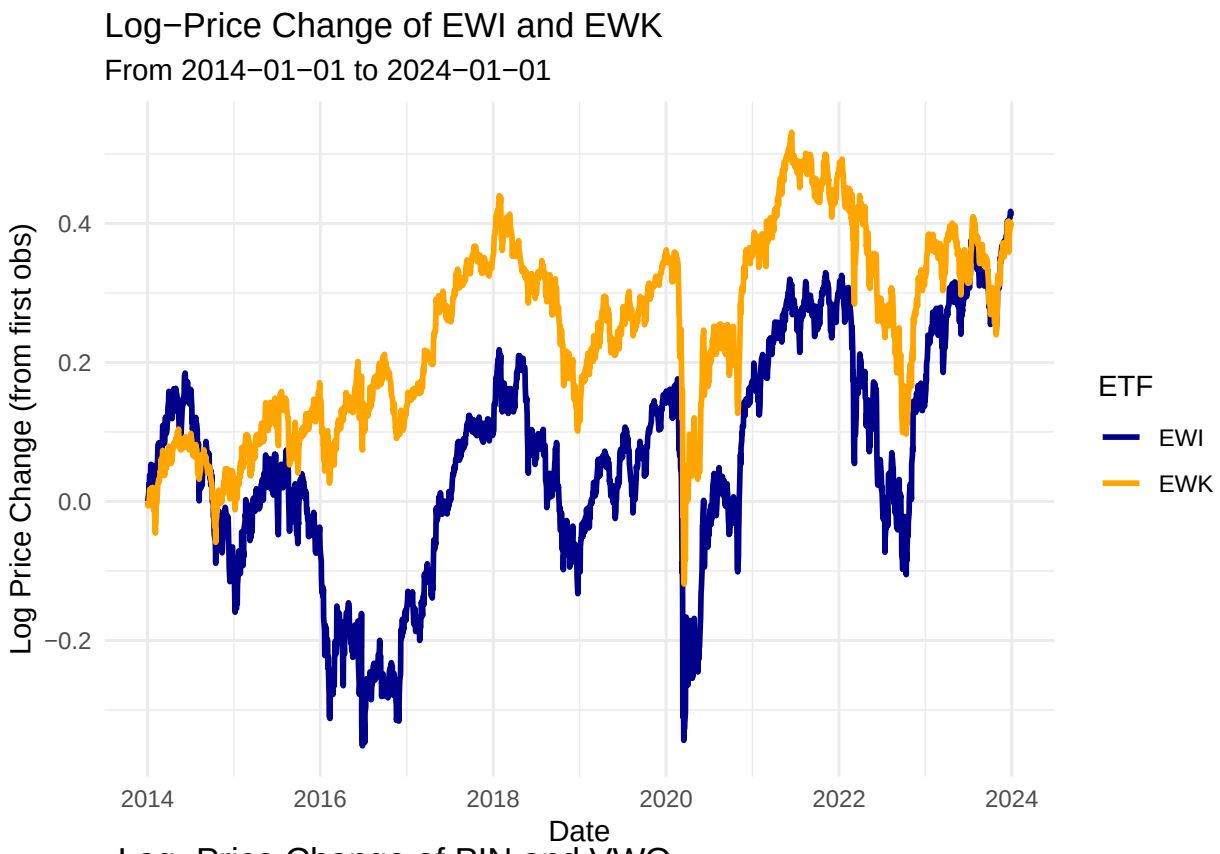

Log−Price Change of EWA and EWL
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
0.50
0.25
0.00
−0.25
2014
2016
2018
2020
2022
2024
Date
ETF
EWA
EWL

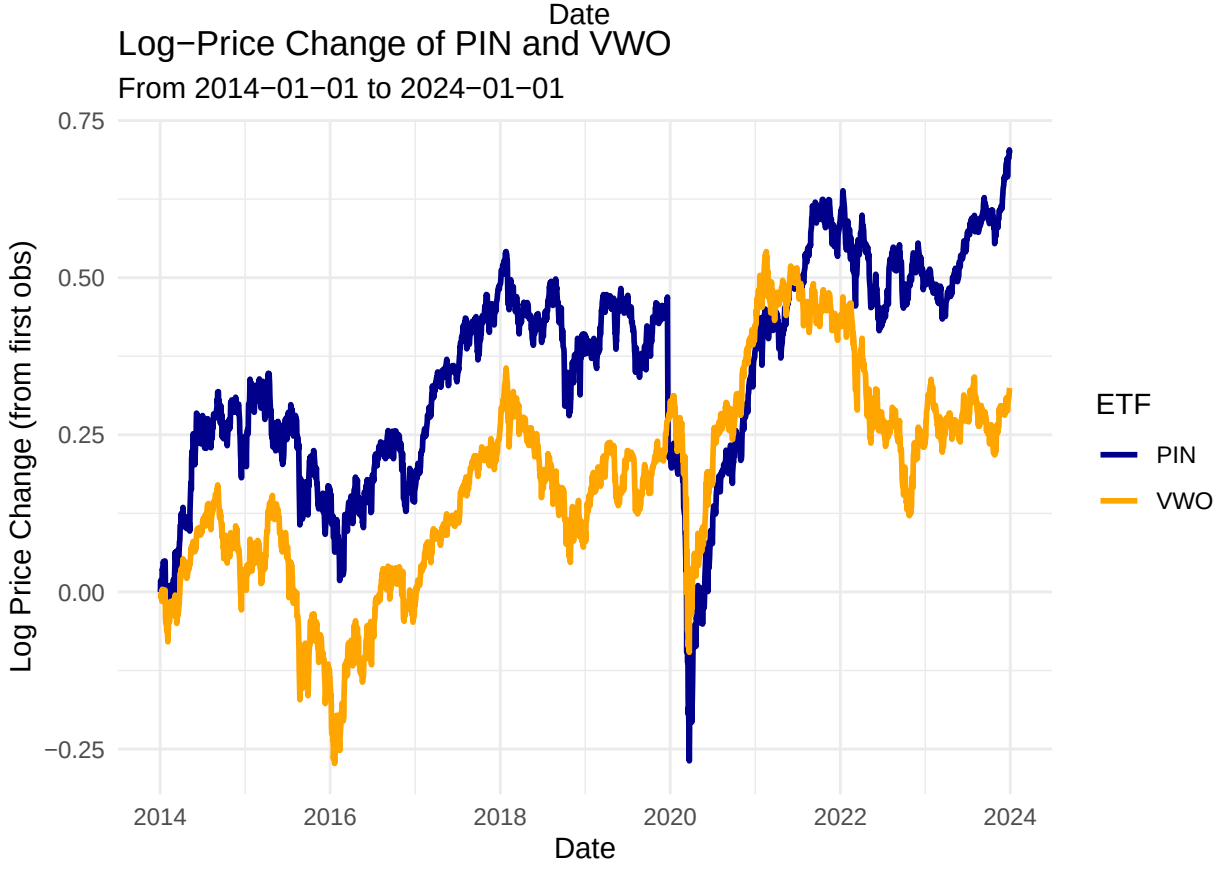

Log−Price Change of EWA and SMH
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
2.0
1.5
1.0
0.5
0.0
2014
2016
2018
2020
2022
2024
Date
ETF
EWA
SMH


Log−Price Change of EWA and SOXX
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
2.0
1.5
1.0
0.5
0.0
2014
2016
2018
2020
2022
2024
Date
ETF
EWA
SOXX


Log−Price Change of EWA and XLB
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
0.5
0.0
2014
2016
2018
2020
2022
2024
Date
ETF
EWA
XLB


Log−Price Change of EWC and EWT
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
0.8
0.4
0.0
−0.4
2014
2016
2018
2020
2022
2024
Date
ETF
EWC
EWT


Log−Price Change of EWI and EWK
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
0.4
0.2
0.0
−0.2
2014
2016
2018
2020
2022
2024
Date
ETF
EWI
EWK


Log−Price Change of PIN and VWO
From 2014−01−01 to 2024−01−01
Log Price Change (from first obs)
0.75
0.50
0.25
0.00
−0.25
2014
2016
2018
2020
2022
2024
Date
ETF
PIN
VWO