Modelling Procedure (ML Fin Data - Project 1)

Hair Albeiro Parra Barrera

Libraries

Getting the data

0.0.1 SP500 Economic Sectors

The following function fetches and extract the economic sectors from the SP500, taken from Wikipedia.

```
# NOTE: not necessary to run anymore
# fetch the sectors as a dataframe
sp500_sectors <- f_get_sp500_sectors()
head(sp500_sectors)
```

```
##   tickers                sectors
## 1     MMM            Industrials
## 2     AOS            Industrials
## 3     ABT            Health Care
## 4    ABBV            Health Care
## 5     ACN Information Technology
## 6    ATVI Communication Services
```

Retrieving top sectors and stocks

The following function will retrieve the top sectors and stocks from the SP500 by weight.

```
# Retrieve top 10 stocks by weight for each sector in the top 5 sectors from the SP500 (by weight)
sector_list <- f_retrieve_top_sp500(top_n_sectors = 6, top_n_stocks = 15, only_tickers=TRUE)
sector_list
```

```
## $Industrials
##  [1] "ADP" "BA"  "CAT" "CSX" "DE"  "ETN" "FDX" "GE"  "HON" "ITW" "LMT" "NOC"
## [13] "RTX" "UNP" "UPS"
##
## $`Health Care`
##  [1] "ABBV" "ABT"  "AMGN" "BMY"  "DHR"  "ELV"  "GILD" "ISRG" "JNJ"  "LLY"
## [11] "MDT"  "MRK"  "PFE"  "TMO"  "UNH"
##
## $`Information Technology`
##  [1] "AAPL" "ACN"  "ADBE" "AMD"  "AVGO" "CRM"  "CSCO" "IBM"  "INTC" "INTU"
## [11] "MSFT" "NVDA" "ORCL" "QCOM" "TXN"
##
## $`Communication Services`
##  [1] "ATVI"  "CHTR"  "CMCSA" "DIS"   "EA"    "GOOG"  "GOOGL" "META"  "NFLX"
## [10] "OMC"   "T"     "TMUS"  "TTWO"  "VZ"    "WBD"
##
## $Financials
```

```
##  [1] "AXP"  "BAC"  "BLK"  "C"    "CB"   "GS"   "JPM"  "MA"   "MMC"  "MS"
## [11] "PGR"  "SCHW" "SPGI" "V"    "WFC"
##
## $`Consumer Discretionary`
##  [1] "ABNB" "AMZN" "AZO"  "BKNG" "CMG"  "F"    "GM"   "HD"   "MAR"  "MCD"
## [11] "NKE"  "ORLY" "SBUX" "TJX"  "TSLA"
```

Retrieving stock data

We will know use the function `f_fetch_all_tickers` under `functions/fetch_sp500_sectors.R`

```
# function to fetch all the information for one ticker into a nice xts dataframe
sp500_stocks <- lapply(sector_list,
                       f_fetch_all_tickers,
                       start_date="2016-01-01",
                       end_date="2022-12-01")
```

The result of this function is a list of lists, with elements as below.

```
# Show the available sectors
names(sp500_stocks)
```

```
## [1] "Industrials"            "Health Care"            "Information Technology"
## [4] "Communication Services" "Financials"             "Consumer Discretionary"
```

```
# Show available stocks for Industrials
names(sp500_stocks$Industrials)
```

```
##  [1] "ADP" "BA"  "CAT" "CSX" "DE"  "ETN" "FDX" "GE"  "HON" "ITW" "LMT" "NOC"
## [13] "RTX" "UNP" "UPS"
```

```
# access the xts of the stocks in industrials
tail(sp500_stocks$Industrials[[1]])
```

```
##            adjusted_close direction_lead discrete_returns realized_returns
## 2022-10-26       230.1928              1      0.008146142      0.009733781
## 2022-11-02       232.4444              1      0.009781309      0.012306110
## 2022-11-09       235.3226              1      0.012382140      0.053616030
## 2022-11-16       248.2840              1      0.055079400      0.034718760
## 2022-11-23       257.0555              1      0.035328500      0.005923399
## 2022-11-30       258.5826             NA      0.005940977               NA
##            adjclose_lag0 adjclose_lag1 adjclose_lag2 adjclose_lag3       atr
## 2022-10-26   0.008113141   0.039930970  -0.064535730   0.030150910  9.676399
## 2022-11-02   0.009733781   0.008113141   0.039930970  -0.064535730  9.885942
## 2022-11-09   0.012306110   0.009733781   0.008113141   0.039930970  9.762661
## 2022-11-16   0.053616030   0.012306110   0.009733781   0.008113141 10.232471
## 2022-11-23   0.034718760   0.053616030   0.012306110   0.009733781 10.243009
## 2022-11-30   0.005923399   0.034718760   0.053616030   0.012306110 10.247795
##                 adx aaron        bb chaikin_vol        clv         emv      macd
## 2022-10-26 13.39493   100 0.6110784 -1.49750300 -0.1320576 -0.01707202 2.049576
## 2022-11-02 13.58997   100 0.6303335  2.90314600 -0.2863719  0.02711271 1.939312
## 2022-11-09 13.77107    50 0.6307783 -0.09676625 -0.3920529  0.04765004 1.866926
## 2022-11-16 14.68326   100 0.8325740 -0.38397100 -0.4461119  0.09074850 1.906715
## 2022-11-23 15.95273   100 0.9310325 -0.20180520 -0.3205142  0.11758529 2.068291
## 2022-11-30 16.53998   100 0.8907336  0.48394890 -0.1089895  0.12144667 2.300754
##                 mfi      sar      smi     volat month_index
## 2022-10-26 51.52422 260.0428 8.131402 0.2269538          82
```

```
## 2022-11-02 49.23300 258.6055  5.546375 0.2606250          83
## 2022-11-09 49.20839 257.2257  3.943960 0.2653165          83
## 2022-11-16 48.83463 256.7200  6.291102 0.2641173          83
## 2022-11-23 49.31528 224.1100 11.099826 0.2624611          83
## 2022-11-30 42.97382 224.1100 16.713518 0.2759187          83
```

BACKTESTING parameters

The following code is used in the `strategy_design.rmd` markdown to simulate the backtesting. You can ignore most of the code here, but some variables are necessary.

```r
# Set up backtesting simulation parameters
sample_xts <- sp500_stocks$Industrials$ADP
sectors <- names(sp500_stocks)
N_sector_best_stocks <- 3 # new strategy: 3x2 = 6

# Formula parameters
slide <- 1
N_months <- length(names(split.xts(sample_xts, f= "months")))
N_window <- 24 # number of months in size for each window
N_runs <- floor((N_months - N_window)/slide)

# display parameters
print(paste0("N_months: ", N_months))
```

```
## [1] "N_months: 83"
```

```r
print(paste0("N_runs: ", N_runs))
```

```
## [1] "N_runs: 59"
```

```r
print(paste0("slide: ", slide))
```

```
## [1] "slide: 1"
```

```r
# setup initial portfolio tracking variables
initial_capital <- 500000
num_tickers <- length(sectors)*N_sector_best_stocks*2 # two sub-strategies for picking
initial_tickers <- rep(NA, num_tickers)
weights <- rep(1/num_tickers, num_tickers) # initialize to 1/n
returns <- rep(NA, N_runs)

# repack the portfolio
portfolio <- list(tickers = initial_tickers,
                  weights = weights,
                  capital = initial_capital,
                  returns = returns,
                  data = NA
                  )
portfolio
```

```
## $tickers
##  [1] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [26] NA NA NA NA NA NA NA NA NA NA NA
##
```

```
## $weights
##  [1] 0.02777778 0.02777778 0.02777778 0.02777778 0.02777778 0.02777778
##  [7] 0.02777778 0.02777778 0.02777778 0.02777778 0.02777778 0.02777778
## [13] 0.02777778 0.02777778 0.02777778 0.02777778 0.02777778 0.02777778
## [19] 0.02777778 0.02777778 0.02777778 0.02777778 0.02777778 0.02777778
## [25] 0.02777778 0.02777778 0.02777778 0.02777778 0.02777778 0.02777778
## [31] 0.02777778 0.02777778 0.02777778 0.02777778 0.02777778 0.02777778
##
## $capital
## [1] 5e+05
##
## $returns
##  [1] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [26] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [51] NA NA NA NA NA NA NA NA NA
##
## $data
## [1] NA
```

MODELLING_PROCEDURE

Recall that the **SECTOR_PROCEDURE**(G, τ) function takes the argument G, which is the **sector name**, and **tau**, which is the current run in the backtesting.

This procedure happens in a loop, for every sector G. Here, we fix one sector only, and a specific τ. The code does the following:

1. Retrieves the actual sector stock data (list of key-value pairs, keys are stock tickers, values are xts full data for that stock.)
2. Creates a variable to store the subset of data that goes into the current window.
3. The `f_extract_window()` function extracts the appropriate window of data corresponding to the τ, with the appropriate window size, for all sectors.
4. Extracts the dynamic features (ARIMA and GARCH) for that each stock in the sector.

```r
# parameters
G <- names(sp500_stocks)[1] # sample sector
tau <- 10 # suppose we are in run 5 of the backtest

####### Inside SECTOR_PROCEDURE ########

# retrieve sector data
sector_data <- sp500_stocks[[G]]

# stocks for sector provided
sector_tickers <- names(sector_data)

# to store subset features for window
sector_stocks_window <- rep(NA, length(sector_tickers))
names(sector_stocks_window) <- sector_tickers

# extract static train-val for all stocks
list_xts_sector <- lapply(sector_data,
                          f_extract_window,
                          tau=tau, # current run
                          n_months = N_window# size of window
                          )

# compute dynamic features for all stocks
```

```
list_xts_sector <- lapply(list_xts_sector,
                          f_extract_dynamic_features,
                          arima_col = "adjusted_close",
                          volat_col = "volat"
                          )

####### Inside SECTOR_PROCEDURE ########

# keys are stock tickers for that sector
names(list_xts_sector)
```

```
##  [1] "ADP" "BA"  "CAT" "CSX" "DE"  "ETN" "FDX" "GE"  "HON" "ITW" "LMT" "NOC"
## [13] "RTX" "UNP" "UPS"
```

```
# each stock has the xts subset (for window)
head(list_xts_sector[[1]])
```

```
##            adjusted_close direction_lead discrete_returns realized_returns
## 2016-10-05       75.58759             -1      0.001486549     -0.008139895
## 2016-10-12       74.97482              1     -0.008106856      0.006425871
## 2016-10-19       75.45815             -1      0.006446562     -0.002748746
## 2016-10-26       75.25101              1     -0.002744971      0.031497430
## 2016-11-02       77.65895              1      0.031998720      0.010172850
## 2016-11-09       78.45300              1      0.010224760      0.025738100
##            adjclose_lag0 adjclose_lag1 adjclose_lag2 adjclose_lag3      atr
## 2016-10-05   0.001485445  -0.016220080   0.024948400  -0.037026570 1.900259
## 2016-10-12  -0.008139895   0.001485445  -0.016220080   0.024948400 1.872384
## 2016-10-19   0.006425871  -0.008139895   0.001485445  -0.016220080 1.800070
## 2016-10-26  -0.002748746   0.006425871  -0.008139895   0.001485445 1.722923
## 2016-11-02   0.031497430  -0.002748746   0.006425871  -0.008139895 1.864142
## 2016-11-09   0.010172850   0.031497430  -0.002748746   0.006425871 1.989560
##                 adx aaron        bb  chaikin_vol         clv           emv
## 2016-10-05 15.44565   -50 0.2934560   -0.4622892  0.18091008 -0.0006643160
## 2016-10-12 15.23639  -100 0.2289285    0.3990933  0.24064338 -0.0026850063
## 2016-10-19 14.75791   -50 0.3060118   -0.4336751  0.09899013 -0.0019094937
## 2016-10-26 14.44363   100 0.2860935   -1.0188680 -0.01496489 -0.0021492280
## 2016-11-02 14.04553    50 0.4910556 -324.8278000  0.05096933 -0.0009225739
## 2016-11-09 13.44222   100 0.5094234    1.1391500  0.19338517 -0.0009562142
##                 macd      mfi      sar        smi      volat month_index
## 2016-10-05 1.3477744 46.50802 95.02127  -5.331162 0.10247324          10
## 2016-10-12 1.1358585 37.92195 94.68802 -11.930732 0.10506831          10
## 2016-10-19 0.9402188 36.19915 94.36810 -17.430099 0.10335977          10
## 2016-10-26 0.7585276 30.28217 94.06097 -19.828752 0.09985285          10
## 2016-11-02 0.6437468 48.88575 93.76613 -18.073978 0.13389984          11
## 2016-11-09 0.5919089 59.37208 93.48309 -13.909935 0.16512456          11
##            sarima_100_001 sarima_010_001 sarima_110_001 sarima_020_001
## 2016-10-05       75.41207       75.25101       75.26197       75.04387
## 2016-10-12       77.80645       77.65895       77.53154       80.06689
## 2016-10-19       78.59602       78.45300       78.41099       79.24705
## 2016-10-26       80.62994       80.49844       80.39021       82.54388
## 2016-11-02       83.53072       83.41565       83.26130       86.33286
## 2016-11-09       82.99005       82.87192       82.90069       82.32819
##            sarima_120_001 sarima_100_011 sarima_010_011 sarima_110_011
## 2016-10-05       75.41138       75.41207       75.25101       75.26197
## 2016-10-12       78.67500       77.80645       77.65895       77.53154
## 2016-10-19       80.10605       78.59602       78.45300       78.41099
## 2016-10-26       81.87782       80.62994       80.49844       80.39021
## 2016-11-02       85.86885       83.53072       83.41565       83.26130
```

```
## 2016-11-09       84.17030       82.99005       82.87192       82.90069
##            sarima_020_011 sarima_120_011 vol_forecast
## 2016-10-05       75.04387       75.41138    0.1338998
## 2016-10-12       80.06689       78.67500    0.1651246
## 2016-10-19       79.24705       80.10605    0.1746223
## 2016-10-26       82.54388       81.87782    0.1752898
## 2016-11-02       86.33286       85.86885    0.1772747
## 2016-11-09       82.32819       84.17030    0.1757262
```

The result is the `list_train_val_sector` object, which is a list of lists. - The first level are the stock tickers - The second level are train and val xts for each stock.

```r
# Check num of rows (weeks) for window
nrow(list_xts_sector[[1]])
```

```
## [1] 103
```

Feature Selection

Notes: - This will use **forward selection** to extract the features from a sample stock for the current sector. - The `target_var` argument specifies the target variable, in this case is called "realized_returns". - `f_select_features()` is found under `functions/feature_engineering.R`

```r
# Extract a sample stock in the list_xts_sector
sample_sector_stock <- list_xts_sector[[1]]

# Define the formula for regression
fmla <- realized_returns ~ . -realized_returns -month_index

# try obtaining best features for a sample train set for a stock in the sample sector
best_feat_list <- f_select_features(
                    fmla = fmla, # formula for regression
                    data = sample_sector_stock, # for one stock of current sector
                    target_var = "realized_returns", # future-lagged log-returns
                    volat_col = "volat", # we always want to keep the volatility col
                    garch_col = "vol_forecast", # GARCH column
                    nvmax = 20, # examine all possible subsets
                    method="exhaustive") #  we always want to use forward selection
```

```
## Loading required package: leaps

## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax, force.in =
## force.in, : 5 linear dependencies found

## Reordering variables and trying again:
```

```r
print("")
```

```
## [1] ""
```

```r
best_feat_list
```

```
## $featnames
##  [1] "adjusted_close" "adjclose_lag1"  "atr"            "emv"
##  [5] "sar"            "volat"          "sarima_100_001" "sarima_010_001"
##  [9] "sarima_020_001" "sarima_120_001" "sarima_110_011" "sarima_020_011"
```

```
## [13] "vol_forecast"
##
## $fmla
## realized_returns ~ adjusted_close + adjclose_lag1 + atr + emv +
##     sar + volat + sarima_100_001 + sarima_010_001 + sarima_020_001 +
##     sarima_120_001 + sarima_110_011 + sarima_020_011 + vol_forecast
## <environment: 0x00000277ca4498c0>
```

The result of this object is a list `best_feat_list` in this case, containing two objects: - `featnames`: a list of features selected. - `fmla`: An R formula (for regression, etc)

NOTE: - This is just for illustration and to visualize the data. The actual feature selection is performed in a loop for every stock as illustrated in the next section. - There will always be linear dependencies because of the ARIMA features. This is normal.

Regularized MLR (Elasticnet)

After feature selection, we want to fit the following model:

$$\mathcal{L}(\beta) = \frac{1}{2}\sum_{i=1}^{n}(y_i - x_i^T\beta)^2 + \lambda\left[\alpha||\beta||_1 + (1-\alpha)||\beta||_2^2\right]$$

First, we wil do the following: 1. Specify the general formula 2. Create the grid of parameters to use in the Elasticnet models 3. Create a tracking variable to save the forecasted returns, MSEs and Sharpe Ratios computed

```r
# load required libraries
library("caret")
library("Metrics")

# Define the formula for regression
fmla <- realized_returns ~ . -realized_returns -month_index

# Create a grid for elastic net regression hyperparameters
grid_enet <- expand.grid(alpha = seq(from = 0, to = 1, by = 0.1),  # Elastic net mixing parameter
                         lambda = seq(from = 0, to = 0.05, by = 0.05))  # Regularization strength

# Initialize variable to save forecasted returns, MSEs and Sharpe Ratios
sector_tracker <- as.list(rep(NA, length(sector_tickers)))
names(sector_tracker) <- sector_tickers

# transform into a list of lists
sector_tracker <- lapply(sector_tracker, function(x) list(
  forecasted_ret = NA,
  sharpe = NA,
  msr = NA, # modified sharpe ratio
  rmse = NA,
  data = NA
))

# display values
fmla # all initial variables
```

```
## realized_returns ~ . - realized_returns - month_index
```

```r
names(sector_tracker) # list of lists
```

```
##  [1] "ADP" "BA"  "CAT" "CSX" "DE"  "ETN" "FDX" "GE"  "HON" "ITW" "LMT" "NOC"
## [13] "RTX" "UNP" "UPS"
```

```r
names(sector_tracker[[1]]) # to store the values as the loop happens
```

```
## [1] "forecasted_ret" "sharpe"         "msr"            "rmse"
## [5] "data"
```

Fitting all the models

Next, we loop through every stock doing the following: 1. Extracting the train and validation sets, and filter NAs 2. Perform feature selection for every stock 3. Fit an Elasticnet model for that stock, and obtain predictions for the returns 4. Compute the RMSE 5. Compute the Sharpe Ratio and Modified Sharpe 6. Save everything

```r
# Loop for every stock ticker in sector G
for(ticker in sector_tickers){
  print(paste0("ticker: ", ticker))

  ### Step 0: Data Preparation

  ###########################################################################
  ### NOTE: Need to refactor

  # fetch data for that ticker
  full_train <- list_xts_sector[[ticker]]

  # Re-extract train and val with full features
  full_train <- f_extract_train_val_no_window(full_train,
                                              val_lag = 1) # number of months in val

  # Reassign to train and val
  ticker_data_train <- full_train$train
  ticker_data_val <- full_train$val

  # remove nas
  ticker_data_train <- na.omit(ticker_data_train) # data cannot contain nas
  ticker_data_val <- na.omit(ticker_data_val) # data cannot contain nas

  ###########################################################################

  ### Step 1: Feature Selection

  # Perform feature selection for that stock
  best_feat_list <- f_select_features(
                      fmla = fmla, # formula for regression
                      data = ticker_data_train, # train data for one stock of current sector
                      target_var = "realized_returns", # forecast future log returns
                      volat_col = "volat", # always keep the actual volatility
                      garch_col = "vol_forecast",
                      nvmax = 20, # total number of max subsets
                      method="exhaustive")

  print(best_feat_list$fmla)

  ### Step 2: Elasticnet

  # Set up time-slice cross-validation parameters
  ctr_train <- trainControl(method = "timeslice", # cross validation
                            initialWindow = 52,   # Consecutive number of weeks
                            horizon = 4,          # Horizon is one month prediction (4 weeks)
                            skip = 1,             # No skip, our data will overlap in practice
```

```
                          fixedWindow = TRUE,   # Use a fixed window
                          allowParallel = TRUE) # Enable parallel processing


# Train the elastic net regression model using time-slice cross-validation
model_enet_best <- train(form = best_feat_list$fmla,           # Formula from feature selection
                         data = ticker_data_train,             # Training data
                         method = "glmnet",                    # Model method = Elasticnet
                         tuneGrid = grid_enet,                 # Hyperparameter grid
                         trControl = ctr_train,                # Cross-validation control
                         preProc = c("center", "scale"),       # Preprocessing steps
                         metric = "Rsquared",                  # Metric for selecting the best model
                         threshold = 0.2)

# Extract the best alpha and beta fitted
best_alpha <- model_enet_best$bestTune$alpha
best_lambda <- model_enet_best$bestTune$lambda

# Use the best-fitted elastic net regression model to make predictions on the val_data
pred_enet_best <- predict(model_enet_best, ticker_data_val) # predict on val
pred_enet_best <- mean(pred_enet_best) # take the average

# Compute the RMSE on the validation set
enet_rmse <- sqrt(mse(actual = ticker_data_val[, "realized_returns"], predicted = pred_enet_best))

### Step 3: Sharpe Ratio

# re-stack train and val
full_train <- rbind.xts(ticker_data_train, ticker_data_val)

# Calculate the Sharpe Ratio and MSR (on historical discrete returns)
scaling_factor <- as.vector(ticker_data_val$month_index)[1] - as.vector(ticker_data_train$month_index)[1]

# Pack returns and compute mean and std
hist_returns <- na.trim(as.vector(full_train[, "discrete_returns"]))
mean_rets <- mean(hist_returns)
std_rets <- sd(hist_returns)

# Calculate the ES and set risk-free
VaR <- quantile(hist_returns, 0.05)
ES <- mean(hist_returns[hist_returns < VaR])
Rf <- 0

# Calculate the Sharpe and MSR
stock_sharpe <- ((mean_rets- Rf)/ std_rets ) * sqrt(scaling_factor) # annualized
stock_msr <- ((mean_rets- Rf)/ ES ) * sqrt(scaling_factor) # annualized

### Step 4: Track the measures

sector_tracker[[ticker]]$forecasted_ret = pred_enet_best
sector_tracker[[ticker]]$rmse = enet_rmse
sector_tracker[[ticker]]$sharpe = stock_sharpe
sector_tracker[[ticker]]$msr = stock_msr
# sector_tracker[[ticker]]$data = rbind.xts(ticker_data_train, ticker_data_val) # This should be included at

# show values
print("******************************************")
print(paste("rmse: ", enet_rmse))
print(paste("sharpe: ", stock_sharpe))
```

```r
  print(paste("msr: ", stock_msr))
  print("******************************************")

  print("###########################################")
}
```

```
## [1] "ticker: ADP"
## Reordering variables and trying again:
## realized_returns ~ adjusted_close + direction_lead + adjclose_lag3 +
##     emv + macd + sar + sarima_100_001 + sarima_010_001 + sarima_020_011 +
##     sarima_120_011 + volat + vol_forecast
## <environment: 0x00000277d4483a68>
## [1] "******************************************"
## [1] "rmse:  0.0760390422855801"
## [1] "sharpe:  1.10177889761402"
## [1] "msr:  -0.443650390427853"
## [1] "******************************************"
## [1] "###########################################"
## [1] "ticker: BA"
## Reordering variables and trying again:
## realized_returns ~ adjusted_close + direction_lead + atr + sar +
##     volat + sarima_110_001 + sarima_120_011 + vol_forecast
## <environment: 0x00000277ddf39af0>
## [1] "******************************************"
## [1] "rmse:  0.0727238024857104"
## [1] "sharpe:  1.70973343797271"
## [1] "msr:  -1.04056567086371"
## [1] "******************************************"
## [1] "###########################################"
## [1] "ticker: CAT"
## Reordering variables and trying again:
## realized_returns ~ discrete_returns + adjclose_lag0 + atr + adx +
##     sar + sarima_010_001 + sarima_110_001 + sarima_020_001 +
##     sarima_120_001 + sarima_110_011 + sarima_020_011 + vol_forecast +
##     volat
## <environment: 0x00000277d9298b18>
## [1] "******************************************"
## [1] "rmse:  0.0432440935220575"
## [1] "sharpe:  0.943881552940442"
## [1] "msr:  -0.507468538506106"
## [1] "******************************************"
## [1] "###########################################"
## [1] "ticker: CSX"
## Reordering variables and trying again:
## realized_returns ~ adjusted_close + direction_lead + discrete_returns +
##     adx + macd + sarima_020_001 + sarima_100_011 + sarima_120_011 +
##     volat + vol_forecast
## <environment: 0x00000277ce2f24e8>
## [1] "******************************************"
## [1] "rmse:  0.0103234105127164"
## [1] "sharpe:  1.1047552756973"
## [1] "msr:  -0.804114655830999"
## [1] "******************************************"
## [1] "###########################################"
## [1] "ticker: DE"
## Reordering variables and trying again:
## realized_returns ~ adjusted_close + discrete_returns + atr +
##     adx + emv + mfi + sar + volat + sarima_010_001 + sarima_110_001 +
##     sarima_120_001 + vol_forecast
```

```
## <environment: 0x00000277d92711f8>
## [1] "******************************************"
## [1] "rmse:  0.0683865227094192"
## [1] "sharpe:  1.02762587088189"
## [1] "msr:  -0.522794467351004"
## [1] "******************************************"
## [1] "##########################################"
## [1] "ticker: ETN"
## Reordering variables and trying again:
## realized_returns ~ adjusted_close + direction_lead + discrete_returns +
##     adjclose_lag0 + aaron + sarima_120_001 + sarima_010_011 +
##     sarima_110_011 + volat + vol_forecast
## <environment: 0x00000277da0bec88>
## [1] "******************************************"
## [1] "rmse:  0.015869391263949"
## [1] "sharpe:  0.740521726447256"
## [1] "msr:  -0.40930849607496"
## [1] "******************************************"
## [1] "##########################################"
## [1] "ticker: FDX"
## Reordering variables and trying again:
## realized_returns ~ adjusted_close + direction_lead + discrete_returns +
##     adjclose_lag2 + adjclose_lag3 + atr + adx + emv + sarima_100_001 +
##     sarima_020_001 + sarima_120_011 + vol_forecast + volat
## <environment: 0x00000277de202d50>
## [1] "******************************************"
## [1] "rmse:  0.0852701861147087"
## [1] "sharpe:  0.673511241109602"
## [1] "msr:  -0.325673333252512"
## [1] "******************************************"
## [1] "##########################################"
## [1] "ticker: GE"
## Reordering variables and trying again:
## realized_returns ~ adjusted_close + direction_lead + adx + bb +
##     chaikin_vol + sar + smi + sarima_110_001 + sarima_020_001 +
##     sarima_120_001 + sarima_010_011 + volat + vol_forecast
## <environment: 0x00000277d842f4e0>
## [1] "******************************************"
## [1] "rmse:  0.162498761379851"
## [1] "sharpe:  -1.20827278549972"
## [1] "msr:  0.475164293908474"
## [1] "******************************************"
## [1] "##########################################"
## [1] "ticker: HON"
## Reordering variables and trying again:
## realized_returns ~ adjusted_close + adjclose_lag0 + adjclose_lag1 +
##     adjclose_lag2 + adjclose_lag3 + atr + adx + bb + macd + mfi +
##     smi + sarima_020_001 + sarima_110_011 + sarima_120_011 +
##     volat + vol_forecast
## <environment: 0x00000277d9561a88>
## [1] "******************************************"
## [1] "rmse:  0.0265816711880368"
## [1] "sharpe:  1.01565170834529"
## [1] "msr:  -0.442514050850984"
## [1] "******************************************"
## [1] "##########################################"
## [1] "ticker: ITW"
## Reordering variables and trying again:
## realized_returns ~ adjusted_close + direction_lead + atr + aaron +
##     mfi + sar + sarima_010_001 + sarima_120_001 + sarima_020_011 +
```

```
##     volat + vol_forecast
## <environment: 0x00000277d114f220>
## [1] "*****************************************"
## [1] "rmse:  0.0229877045333249"
## [1] "sharpe:  0.465646659189681"
## [1] "msr:  -0.200112872540506"
## [1] "*****************************************"
## [1] "#########################################"
## [1] "ticker: LMT"
## Reordering variables and trying again:
## realized_returns ~ adjusted_close + direction_lead + adjclose_lag0 +
##     adjclose_lag2 + adx + bb + chaikin_vol + clv + sar + sarima_110_001 +
##     sarima_020_001 + sarima_120_001 + vol_forecast + volat
## <environment: 0x00000277d9908bc0>
## [1] "*****************************************"
## [1] "rmse:  0.0952911862310501"
## [1] "sharpe:  0.800549448568796"
## [1] "msr:  -0.396725888105181"
## [1] "*****************************************"
## [1] "#########################################"
## [1] "ticker: NOC"
## Reordering variables and trying again:
## realized_returns ~ adjusted_close + direction_lead + adjclose_lag1 +
##     adx + aaron + smi + volat + sarima_120_001 + sarima_100_011 +
##     sarima_020_011 + vol_forecast
## <environment: 0x00000277de78d5a8>
## [1] "*****************************************"
## [1] "rmse:  0.0732283905281539"
## [1] "sharpe:  0.765604048283122"
## [1] "msr:  -0.320178164195755"
## [1] "*****************************************"
## [1] "#########################################"
## [1] "ticker: RTX"
## Reordering variables and trying again:
## realized_returns ~ adjusted_close + adjclose_lag1 + adjclose_lag3 +
##     adx + clv + macd + sar + smi + sarima_100_001 + sarima_120_001 +
##     sarima_110_011 + sarima_120_011 + vol_forecast + volat
## <environment: 0x00000277d07f2948>
## [1] "*****************************************"
## [1] "rmse:  0.0692210584201641"
## [1] "sharpe:  0.848488763173833"
## [1] "msr:  -0.412356870343423"
## [1] "*****************************************"
## [1] "#########################################"
## [1] "ticker: UNP"
## Reordering variables and trying again:
## realized_returns ~ adjusted_close + adx + aaron + bb + chaikin_vol +
##     emv + smi + sarima_110_011 + sarima_120_011 + volat + vol_forecast
## <environment: 0x00000277db76a810>
## [1] "*****************************************"
## [1] "rmse:  0.0822333759466835"
## [1] "sharpe:  1.06425677754482"
## [1] "msr:  -0.574006467179498"
## [1] "*****************************************"
## [1] "#########################################"
## [1] "ticker: UPS"
## Reordering variables and trying again:
## realized_returns ~ adjusted_close + direction_lead + discrete_returns +
##     adjclose_lag0 + adjclose_lag2 + atr + bb + chaikin_vol +
##     clv + macd + mfi + smi + volat + sarima_010_001 + sarima_110_001 +
```

```
##     sarima_120_011 + vol_forecast
## <environment: 0x00000277d9d26378>
## [1] "******************************************"
## [1] "rmse:  0.112930518581764"
## [1] "sharpe:  0.283233321392086"
## [1] "msr:  -0.115446797219559"
## [1] "******************************************"
## [1] "##########################################"
```

Now that all the models have been trained and the metrics recorded, we now simply choose the top 3 stocks based on the return, and the top 3 based on the best sharpe or modified sharpe ratio.

Let's first show some values for the `sector_tracker` object:

```r
names(sector_tracker)
```

```
##  [1] "ADP" "BA"  "CAT" "CSX" "DE"  "ETN" "FDX" "GE"  "HON" "ITW" "LMT" "NOC"
## [13] "RTX" "UNP" "UPS"
```

```r
names(sector_tracker[[1]])
```

```
## [1] "forecasted_ret" "sharpe"         "msr"            "rmse"
## [5] "data"
```

```r
sector_tracker
```

```
## $ADP
## $ADP$forecasted_ret
## [1] -0.06550883
##
## $ADP$sharpe
## [1] 1.101779
##
## $ADP$msr
## [1] -0.4436504
##
## $ADP$rmse
## [1] 0.07603904
##
## $ADP$data
## [1] NA
##
##
## $BA
## $BA$forecasted_ret
## [1] -0.03679629
##
## $BA$sharpe
## [1] 1.709733
##
## $BA$msr
## [1] -1.040566
##
## $BA$rmse
## [1] 0.0727238
##
## $BA$data
## [1] NA
```

```
##
##
## $CAT
## $CAT$forecasted_ret
## [1] -0.009120676
##
## $CAT$sharpe
## [1] 0.9438816
##
## $CAT$msr
## [1] -0.5074685
##
## $CAT$rmse
## [1] 0.04324409
##
## $CAT$data
## [1] NA
##
##
## $CSX
## $CSX$forecasted_ret
## [1] -0.005156274
##
## $CSX$sharpe
## [1] 1.104755
##
## $CSX$msr
## [1] -0.8041147
##
## $CSX$rmse
## [1] 0.01032341
##
## $CSX$data
## [1] NA
##
##
## $DE
## $DE$forecasted_ret
## [1] -0.04339525
##
## $DE$sharpe
## [1] 1.027626
##
## $DE$msr
## [1] -0.5227945
##
## $DE$rmse
## [1] 0.06838652
##
## $DE$data
## [1] NA
##
##
## $ETN
## $ETN$forecasted_ret
## [1] 0.0008634067
##
## $ETN$sharpe
## [1] 0.7405217
##
```

```
## $ETN$msr
## [1] -0.4093085
##
## $ETN$rmse
## [1] 0.01586939
##
## $ETN$data
## [1] NA
##
##
## $FDX
## $FDX$forecasted_ret
## [1] 0.07756788
##
## $FDX$sharpe
## [1] 0.6735112
##
## $FDX$msr
## [1] -0.3256733
##
## $FDX$rmse
## [1] 0.08527019
##
## $FDX$data
## [1] NA
##
##
## $GE
## $GE$forecasted_ret
## [1] 0.1436685
##
## $GE$sharpe
## [1] -1.208273
##
## $GE$msr
## [1] 0.4751643
##
## $GE$rmse
## [1] 0.1624988
##
## $GE$data
## [1] NA
##
##
## $HON
## $HON$forecasted_ret
## [1] -0.0178617
##
## $HON$sharpe
## [1] 1.015652
##
## $HON$msr
## [1] -0.4425141
##
## $HON$rmse
## [1] 0.02658167
##
## $HON$data
## [1] NA
##
```

```
##
## $ITW
## $ITW$forecasted_ret
## [1] 0.01226767
##
## $ITW$sharpe
## [1] 0.4656467
##
## $ITW$msr
## [1] -0.2001129
##
## $ITW$rmse
## [1] 0.0229877
##
## $ITW$data
## [1] NA
##
##
## $LMT
## $LMT$forecasted_ret
## [1] -0.07485236
##
## $LMT$sharpe
## [1] 0.8005494
##
## $LMT$msr
## [1] -0.3967259
##
## $LMT$rmse
## [1] 0.09529119
##
## $LMT$data
## [1] NA
##
##
## $NOC
## $NOC$forecasted_ret
## [1] -0.05570425
##
## $NOC$sharpe
## [1] 0.765604
##
## $NOC$msr
## [1] -0.3201782
##
## $NOC$rmse
## [1] 0.07322839
##
## $NOC$data
## [1] NA
##
##
## $RTX
## $RTX$forecasted_ret
## [1] -0.05114387
##
## $RTX$sharpe
## [1] 0.8484888
##
## $RTX$msr
```

```
## [1] -0.4123569
##
## $RTX$rmse
## [1] 0.06922106
##
## $RTX$data
## [1] NA
##
##
## $UNP
## $UNP$forecasted_ret
## [1] -0.06689051
##
## $UNP$sharpe
## [1] 1.064257
##
## $UNP$msr
## [1] -0.5740065
##
## $UNP$rmse
## [1] 0.08223338
##
## $UNP$data
## [1] NA
##
##
## $UPS
## $UPS$forecasted_ret
## [1] 0.09726426
##
## $UPS$sharpe
## [1] 0.2832333
##
## $UPS$msr
## [1] -0.1154468
##
## $UPS$rmse
## [1] 0.1129305
##
## $UPS$data
## [1] NA
```

```r
# Extract the top 3 tickers with the highest Sharpe ratio
top_sharpe <- names(sort(sapply(sector_tracker, function(x) x$sharpe), decreasing=TRUE))[1:3]
top_fore_rets <- names(sort(sapply(sector_tracker, function(x) x$forecasted_ret), decreasing=TRUE))[1:3]

# display selected stocks
top_fore_rets
```

```
## [1] "GE"  "UPS" "FDX"
```

```r
top_sharpe
```

```
## [1] "BA"  "CSX" "ADP"
```

```r
## TODO: Complete the function, keep the name and parameters
f_select_top_stocks <- function(sector_tracker, n=3){
  ## selects the top n + n stocks (n based on forecasted return, n based on sharpe)
```

```r
  ##
  ## Params:
  ##    - sector_tracker (list of lists): generated by the Loop for every stock ticker in sector G
  ##    - n (int): number of top stocks to choos efor each method. Top n for the predicted returns,
  ##               and top n for the sharpe-based.

  # Extract the top 3 tickers with the highest Sharpe ratio
  top_sharpe <- names(sort(sapply(sector_tracker, function(x) x$sharpe), decreasing=TRUE))[1:n]
  top_fore_rets <- names(sort(sapply(sector_tracker, function(x) x$forecasted_ret), decreasing=TRUE))[1:n]

  # Concat in one list
  top_tickers  <-c(top_sharpe, top_fore_rets)

  # Create a new named list with tickers and their corresponding data
  best_stocks_data <- lapply(top_tickers, function(x) sector_tracker[[x]])
  names(best_stocks_data) <- top_tickers

  return(best_stocks_data)
}
```

```r
# Cbtain data for the top n*2 stocks (best forecasted rets and best sharpe together)
f_select_top_stocks(sector_tracker = sector_tracker, n = 2)
```

```
## $BA
## $BA$forecasted_ret
## [1] -0.03679629
##
## $BA$sharpe
## [1] 1.709733
##
## $BA$msr
## [1] -1.040566
##
## $BA$rmse
## [1] 0.0727238
##
## $BA$data
## [1] NA
##
##
## $CSX
## $CSX$forecasted_ret
## [1] -0.005156274
##
## $CSX$sharpe
## [1] 1.104755
##
## $CSX$msr
## [1] -0.8041147
##
## $CSX$rmse
## [1] 0.01032341
##
## $CSX$data
## [1] NA
##
##
## $GE
## $GE$forecasted_ret
```

```
## [1] 0.1436685
##
## $GE$sharpe
## [1] -1.208273
##
## $GE$msr
## [1] 0.4751643
##
## $GE$rmse
## [1] 0.1624988
##
## $GE$data
## [1] NA
##
##
## $UPS
## $UPS$forecasted_ret
## [1] 0.09726426
##
## $UPS$sharpe
## [1] 0.2832333
##
## $UPS$msr
## [1] -0.1154468
##
## $UPS$rmse
## [1] 0.1129305
##
## $UPS$data
## [1] NA
```