BTMA 431 (Fall 2022) HW #1: Exploring Financial Data

Purpose of This Assignment

In this assignment, you will do some basic exploratory data analysis with financial data. You will create new columns and data frames based on previously defined ones, and you will create plots in R.

Instructions

The submission instructions and grading criteria are on the last page of this pdf. Please read them carefully.

Disclaimer

None of this is to be taken as financial advice. It is not recommended that you put all your money into an asset class based off of a homework assignment. Also, none of the scenarios in this assignment or in this class are based on real people. If you happen to know someone who fits a scenario, then that is just a coincidence.

Exploring Financial Data

Your friend has just learned that his long-lost relative, who passed away years ago, had named him as a beneficiary in a will that was just found. His Great-Aunt Belinda was an original cypherpunk in the 1980s. Shortly before her passing, she bought $10,000 worth of Bitcoin (BTC) on December 1, 2017. She also created a portfolio with $5,000 in Ripple (XRP) and $5,000 in Litecoin (LTC). She created another portfolio which consisted of $2,500 in Ethereum (ETH), $2,500 in Dash (DASH), $2,500 in Peercoin (PPC), and $2,500 in Stellar (XLM). These were all in the top 10 cryptocurrencies by market capitalization around the time when she made these purchases. In her will, she says that your friend will have first dibs on one of these portfolios, and his distant cousins who he has never met would have the leftovers. He worries that if he takes too long to decide, then the cousins will try to fight him in court to contest the will. He wants your advice. However, before you make a recommendation, you want to understand these portfolios better by examining the historical price data.

For the remainder of the problem, define the above three portfolios as follows: Portfolio 1: 100% BTC; Portfolio 2: 50% XRP and 50% LTC; Portfolio 3: 25% ETH, 25% DASH, 25% PPC, and 25% XLM. The weights (the percentages) are with respect to the prices on December 1, 2017.

Note: As a simplification, **only use closing price data** from December 1, 2017 to September 1, 2022 (**including both of those dates**) for all of sub-questions of Questions 1 and 2. The closing price is in the 'close' column for each dataset. Also assume that you can buy fractional amounts (i.e., no need to find the max number of BTC below some fractional amount of BTC). Ignore any *airdrops* or *staking rewards* these cryptocurrencies may have had (if any of these have had any *hard forks* or migrated to a *proof-of-stake consensus algorithm*). Also, every dollar amount (both in the assignment statement and in the dataset) is in terms of USD (US dollars). Your friend is living in the States and cares about the portfolio in USD.

Tip: It will be useful for you to think at a high level about out how to make a for-loop for the first three sub-questions (so that if instead of seven, if I had asked you about twenty cryptocurrencies or twenty stocks, you would be able to generalize your approach without much extra effort). However, we may not have learned the functions that would help us write a for-loop to solve the problem yet, so don't worry about writing one if you cannot figure it out. We will see more for-loops in the coming weeks.

Before you begin, load the data by saving the 'HW1_data2022.rData' file from the HW 1 D2L folder to your working directory. If you do not know where your working directory folder is, then use the getwd() function. If you want to set your working directory to another folder, then you can use setwd():

```
setwd("C:/Documents/Teaching/Fall 2022/HW") # Modify this to your folder
```

To load the data, use the load() function:

```
load("HW1_data2022.rData") # Loads data from the working directory into the current session
```

The data was obtained from Yahoo Finance using the *quantmod* package. The same data could be downloaded from CoinMarketCap (https://coinmarketcap.com/) using their API, however, we won't focus on retrieving the data in this assignment; you will learn how to gather data from the web later.

Now that we have the data, we can begin to explore the data.

Starting Simple: Focusing on Individual Cryptocurrencies

1a. (10 points) Of the seven cryptocurrencies, which has had the highest long-run ROI (defined below)?

$$\text{Long-Run ROI} = \frac{\text{value on Sep 1, 2022} - \text{value on Dec 1, 2017}}{\text{value on Dec 1, 2017}}.$$

1b. (10 points) Which one has the highest mean daily return?

1c. (10 points) Which one has the lowest standard deviation of its daily return (i.e., lowest risk)?

Expanding the Exploration: Examining the Portfolios

2a. (10 points) This is likely the question your friend is most interested in if he plans on cashing out now: How much is each portfolio is worth? Find out how much each portfolio was worth on September 1, 2022 (we'll use the same end date so that everyone gets the same answer on the D2L quiz associated to HW 1).

2b. (15 points) Plot the value of the three portfolios over time, either as a combined plot or as separate plots. What was the highest value achieved by each of the three portfolios between December 1, 2017 and September 1, 2022? Among the three portfolios, which portfolio achieved the highest value?

2c. (10 points) How correlated is Portfolio 2's daily return with Bitcoin's daily return? How correlated is Portfolio 3's daily return with Bitcoin's daily return? Plot two separate scatterplots for the pairwise daily returns, and then find the correlation coefficients using the cor() function.

2d. (10 points) Your friend is also curious about the volatility of each portfolio. Notice that to get volatility, we calculate the standard deviation of the daily returns. Among the three portfolios, which one had the highest volatility between December 1, 2017, and December 31, 2019? Did the same portfolio have the highest volatility after the Covid outbreak (from January 1, 2020, to September 1, 2022)

2e. (20 points) For Portfolios 2 and 3, split the data into two groups: 1) Observations on days for which Bitcoin's daily return was positive, 2) Observations on days for which Bitcoin's daily return was negative. What was the mean daily return for Portfolio 2 when focusing on the days when Bitcoin's daily return was positive? What was the mean daily return for Portfolio 2 when focusing on the days when Bitcoin's daily return was negative? Similarly, what was the mean daily return for Portfolio 3 when focusing on the days when Bitcoin's daily return was positive? What was the mean daily return for Portfolio 3 when focusing on the days when Bitcoin's daily return was negative?

Specifically, did one portfolio dominate the other (when assessed and ranked in terms of mean daily return) both when BTC went up and when BTC went down? If so, which one? Alternatively, was it the case that the portfolio that did better on average when BTC went up also did worse on average when BTC went down?

Robustness Checks

3. (5 points) One question you might wonder is how robust your responses are to slight changes in the cryptocurrency prices. Suppose that instead of the Closing price for each day, you had used the average between the High and Low prices for that day for your daily return and long-run ROI calculations. Of the previous questions, which of your responses would change? As is the case for the other questions on this assignment, you must have code to justify your response to get credit for this question.

Instructions

First, read all of the instructions and questions carefully for the entire assignment.

Then for each sub-problem, before writing any code, think about the problem logically. Understand what data is available, so that you can think of ways to use that data to create what you want to create. Write down a structured, logical solution approach in words (in plain English, without code). Your approach should be detailed enough so that someone else would be able to implement your approach in code. **Write your approach as R comments before your solution for that problem or sub-problem**. Note that this is an iterative process. As you are solving the problem, if you realize that your approach was fundamentally missing something, then you should revise your solution approach above.

Grading Criteria

I will mark the assignment in the following way: I will look at the D2L score of your final D2L HW quiz attempt and take off points if your code does not work. **You must submit your answers to the questions in the D2L quiz for the homework to get credit for the assignment**. Your R file must be able to execute properly and replicate your D2L quiz responses to get credit. If your R code has a bug anywhere (if I try to run a line of code and it doesn't work), then you will lose points. If you don't submit an R file, if you don't submit a HW 1 D2L Quiz attempt, or if your R code and plots do not match your responses, then you will get zero credit.

Submit your R code as **a single R file** to the D2L dropbox folder. In the file name, include the homework number, your first name, your last name, and your section number. An example would be 'HW1_firstName_lastName_L02.R' or 'HW1_firstName_lastName_L02.Rmd' (depending on whether or not you used RMarkdown).

Caution: Early on in the semester, students do not follow instructions carefully. If you follow the steps below, you can almost guarantee that you will not lose points due to your R code not running properly on my computer. Carefully making sure your code works fine and does not contain irrelevant chunks of code is good practice, and you want to have formed these habits prior to working together in teams on projects.

1) Save your current homework R file, and close RStudio.
2) Open your homework R file again.
3) Start at Line 1 (the first line in your code). Press Ctrl + Enter (Cmd + Enter for Macs).
4) Line by line, keep pressing Ctrl + Enter (Cmd + Enter) and check the outputs. Make sure that your outputs match your D2L quiz responses for the assignment.
5) Remove all extraneous bits or chunks of code that were not needed to produce your output. For example, if you had bits of code that were dead ends and the objects you created were not used in your solution, then delete those lines of code. Students have things like 'Attempt 4' or 'Attempt 10' in their code. Before doing this step, you may want to save a separate copy that contains all your dead ends (in case you accidentally delete something you needed). That's great to have a separate copy of your work (saved as a separate R file) containing the dead ends, but in your final submission, only submit code that is part of your working solution. Otherwise, if you include all prior attempts before your working solution, then the TA might accidentally mark the wrong approach and flag your work for potential academic misconduct (in which case, I would have to spend time to investigate).
6) **Repeat Steps 1 - 5 again (to make sure you didn't remove a needed line of code).**
7) Save your file with your name as stated in the instructions, and submit your work to D2L. Your work for the assignment should all be in a single R file, not multiple R files and not a zip file.