Project 2 Proposal: Determine Top Performing Crypto Wallets

Name: Chun Him Cheung, Dominic Lim, Ethan Nguonly, Sarah Zheng
GitHub repository: https://github.com/UC-Berkeley-I-School/Project2_Cheung_Lim_Nguonly_Zheng
Primary dataset: https://ethplorer.io/ or https://api.ethplorer.io/ to track address portfolios (i.e This is a sample dataset for wallet portfolio: https://ethplorer.io/address/0xb9c764114c5619a95d7f232594e3b8dddf95b9cf). In order to find the top-performing wallets, we'll have to pull data for multiple wallet portfolios as part of our project.

Introduction

We want to determine what are the top performing Ethereum wallets for us to track, in order to gain an investment edge over Ethereum's market rate of return. Actively traded wallets, where traders may have a deeper knowledge of Ethereum-based coins, may have a significantly higher rate of return vs. passively holding Ether (ETH), the native token of the Ethereum blockchain. Because wallets can hold various different coins, we will take a two-step analysis, where we first identify the highest performing wallets, and then we analyze the transaction history of individual coins in these wallets to determine if any particular coins disproportionately contributed to each wallet's success.

Initial plots, figures, or tables

Our initial exploratory data analysis will begin with tracking the profit and loss of the top 100 wallet addresses. We will need to first prime the dataset by removing gifted tokens as a result of monetary policy to explore purchases and sales only. This will ensure that we are evaluating the "active management" of these address holders. In addition to tracking total returns, we would also like to track the variability (i.e risk) of those returns. By overlaying the "alpha" generated, defined as excess returns over ETH (or a benchmark cryptocurrency) in that time period, our plots would help to visualize the best wallet holders and addresses to track.

The second stage requires an attribution analysis in which we track the profit and variability of those returns by cryptocurrency. Here, we can produce two visuals: a chart that shows the total return of the wallets on one axis and the "risk-adjusted" return on another axis, and another chart that aggregates the total earnings across all wallets and shows how individual coins contributed to the overall earnings.

Variables to explore and expected insights

Given our report is divided into two main areas of focus -- top-performing wallets and top-performing coins in those wallets, respectively -- the variables we plan on exploring will differ depending on which phase we are on in our research and may require supplemental datasets. For baseline reference regarding our first research question, these are the initial columns we'll likely use in our dataset to properly identify which are the top performing wallets.

- Wallet address
- Date of transaction

- fromAddress and toAddress
- tokenSymbol
- Value transferred / Value in USD / Change in value
- Balance before / after transfer
- Wallet top holdings
- Return of Ethereum to use as a benchmark

Once we have identified the top-performing wallet, this is likely where we'll need supplemental datasets to pull the real-time/past price performance and combine it with the wallet's crypto holdings data we've already explored. This way, we'll be able to join the price data with the wallet's transaction history to determine the cryptocurrency's USD price at each timestamp, and ultimately map out the crypto's performance. The crypto with the highest positive impact on the wallet's overall performance is the information we are seeking to understand. With the result of this data, we'll be able to take these insights on which cryptos are the highest performing and apply them to our own investments to yield a higher return.

What you plan on covering in the final report and how you plan to organize
As mentioned above, in the final report, we plan to identify the wallets among the 100 largest Ethereum wallets which showed the highest degree of 'alpha' in terms of making investment decisions and outperforming the passive strategy. We will also conduct some return attribution to show which specific coins within those wallets were responsible for producing that alpha. Lastly, we will also apply principles from portfolio theory and analysis of active portfolio management to break down the source of alpha. For example, was alpha generation concentrated in a small number of well-timed trades or was the alpha evenly distributed over our sample period?

These findings are useful for two general purposes. Firstly, by identifying the top-performing wallets, we can track movements of coins into and from those specific wallets to give us an investment signal on when to make a cryptocurrency investment. The assumption behind this is that wallets which have traditionally outperformed the passive index are likely to belong to insiders in the cryptocurrency community (i.e. the whales) who may be privy to knowledge that others don't have. Secondly, by examining the investment style of those who have traditionally "beat the market", we can help identify optimal investment strategies for Ethereum-based cryptocurrencies. This second purpose can help us generate a generalized investment framework on when and how to invest in crypto-projects given their current stage of market development in the Ethereum network.