A Kernel Density Based Approach to Portfolio Optimization

Lawrence Liu

June 2023

Introduction: Problem Statement

Goal: Portfolio Optimization, given a set of m stocks, find the allocation of funds to each stock that maximizes the sharpe ratio:

$$\text{Sharpe Ratio} = \frac{\mu_r}{\sigma_r}\sqrt{252} \tag{1}$$

Where μ_r is the expected day to day return of the portfolio, σ_r is the standard deviation of the day to day returns of the portfolio.
Constraints: The sum of the weights must be 1, and each weight must be between 0 and 1. And we must use a buy and hold strategy, i.e. we cannot change the amount of money allocated to each stock over time.

Introduction: Modern Portfolio Theory

- Previous approach: **Markowitz Portfolio Optimization**
- Assume the returns of each stock are normally distributed, ie if the day to day return of all the stocks r distributed as $r \sim \mathcal{N}(\mu, \Sigma)$
- Then the expected return of the portfolio is $\mu_r = w^T \mu$ and the variance of the portfolio is $\sigma_r^2 = w^T \Sigma w$

$$
\begin{aligned}
\underset{w}{\text{maximize}} \quad & \frac{w^T \mu}{\sqrt{w^T \Sigma w}} \\
\text{subject to} \quad & 1^T w_i = 1 \\
& w_i \geq 0 \quad \forall i \in \{1, \ldots, n\}
\end{aligned} \tag{2}
$$

Introduction: Solution to Fractional Programming

To solve that equation we do the following.
We introduce a dummy variable $y = \alpha w$, where $\alpha = 1^T y$. Then we can reformulate the problem as:

$$\begin{aligned} \underset{y}{\text{minimize}} \quad & y^T \Sigma y \\ \text{subject to} \quad & \mu^T y = 1 \\ & y_i \geq 0 \quad \forall i \in \{1, \ldots, n\} \end{aligned} \tag{3}$$

This is a convex problem and can be solved using a quadratic programming solver. Then we have $w = \frac{y}{1^T y} = \frac{y}{\alpha}$.

Introdcution: Problems with Modern Portfolio Theory

MPT makes the following assumptions:

- The returns of stocks are not normally distributed
- The returns of the stocks are stationary with time

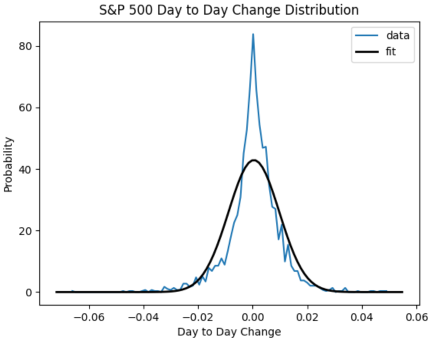
Today we will show that the first assumption is not true and propose a new method that is not reliant on the first assumption

Methods: Data

- We scraped the data from Yahoo Finance using the `yfinance` python package for all the stocks in the S&P 500
- We used the data from 2010-01-01 to 2020-01-01
- Because some tickers were not trading for the entire time period, we only used the tickers that were trading for the entire time period, thus we had 428 companies in our dataset.
- We separated the data into the first 8 years being train and the last 2 years being test we used to evaluate our model on.

Non-Normality of Stock Returns

- Perform a Kolmogorov-Smirnov Test on the day to day changes in stock price for each ticker
- At Significance 0.001, we can reject the Null Hypothesis that the day to day changes in stock price are distributed according a fitted Normal Distribution
- Also plotted out distribution of day to day change of the S&P 500 stock price and fitted normal



Our Method

Our method is twofold:

- A nonparametric Kernel Density Estimation (KDE) to obtain a more accurate estimate for the portfolio variance
- A spectral clustering based method to break the problem into smaller sectors to first find an optimal portfolio for

Kernel Desnity Estimation

We estimate the multivariate distribution of all day to day changes of all the stocks x in our dataset using a kernel density estimation.

$$\hat{f}_{\Sigma}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{\Sigma}(x - x_i') \tag{4}$$

Where x_i' is the ith training point. In our case x_i' is a m dimensional vector representing the daily change for each of the m stocks in our dataset. We use a Gaussian Kernel

$$K_{\Sigma}(x) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x\right) \tag{5}$$

Kernel Density Estimation: Optimizing Σ

We optimize Σ by splitting the training data into a training and validation set. We then optimize Σ by minimizing the negative log likelihood of the validation set.

$$\mathcal{L}(x_1, \ldots, x_k) = -\sum_{i=1}^{k} \log(\hat{f}_\Sigma(x_i)) \tag{6}$$

Where $x_1, \ldots, x_k$ are the validation datapoints.
Because Σ is positive semidefinite, we use Cholesky factorization $\Sigma = R^T R$ to optimize R instead.

$$\frac{\partial}{\partial R}\mathcal{L}(x_1, \ldots, x_k) \approx \sum_{i=1}^{k} \frac{1}{\hat{f}(x_i)} \sum_{j=1}^{n} K_\theta(x_i - x_j') \Big(R^{-T} - \Sigma^{-1}(x_i - x_j')(x_i - x_j')^T R^{-1} \Big)$$

Kernel Density Estimation: A More Accuracte Estimate of Portfolio Variance

From some math we get that the portfolio variance is given by

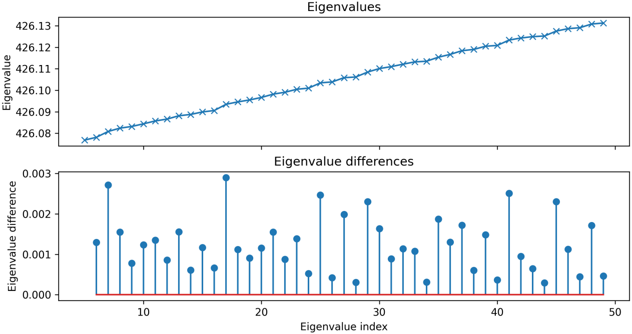
$$\mathbb{E}[(w^T x)^2] - \mathbb{E}^2[w^T x] = w^T \left(\Sigma + \frac{1}{n} \sum_{i=1}^{n} x_i' x_i'^T - \left(\frac{1}{n} \sum_{i=1}^{n} x_i' \right) \left(\frac{1}{n} \sum_{i=1}^{n} x_i' \right)^T \right) w \quad (7)$$

Therefore we can isolate the part between the w^T and w as the covariance matrix of the kernel density estimate.

Spectral Clustering

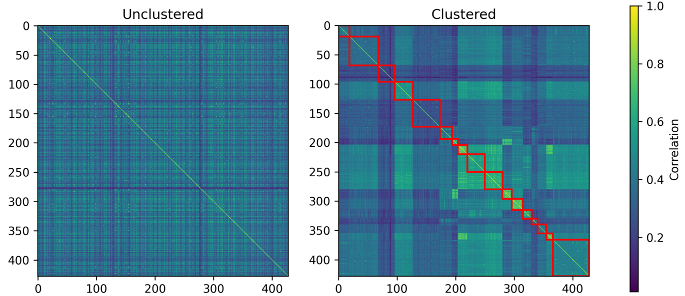
- To identify each sector, we would want stocks that are highly correlated with each other to be in the same sector
- Thus we use spectral clustering, with the adjacency matrix being the negative of the correlation matrix of the stocks
- To identify the number of sectors, we use the eigengap heuristic, but we limit the number of sectors to be between 5 and 50

Spectral Clustering Results: Eigengap



We used 17 sectors for our model

Spectral Clustering Results: Correlation Matrix



Spectral Clustering Results: Individual Sectors

Ticker	Company
AMP	Ameriprise Financial, Inc.
BK	The Bank of New York Mellon Corporation
C	Citigroup Inc.
FITB	Fifth Third Bancorp
JPM	JPMorgan Chase & Co.
LNC	Lincoln National Corporation
MET	MetLife, Inc.
NTRS	Northern Trust Corporation
PNC	The PNC Financial Services Group, Inc.
PRU	Prudential Financial, Inc.
RJF	Raymond James Financial, Inc.
STT	State Street Corporation
HIG	The Hartford Financial Services Group, Inc.
TFC	Truist Financial Corporation
USB	U.S. Bancorp
WFC	Wells Fargo & Company

Overall Results

Method	Sharpe Ratio (train)	Sharpe Ratio (test)
S&P 500	0.804	0.709
Markowitz	2.266	1.199
KDE	1.201	0.920
Spectral Clustering + KDE	1.969	1.424

Method	YOY (train)	YOY (test)
S&P 500	11.352 %	9.947%
Markowitz	37.699%	15.969%
KDE	19.362%	12.743%
Spectral Clustering + KDE	25.246%	17.398%

Discussion and Conclusion

- Spectral Clustering is able to cluster the correlation matrix to be more block diagonal, and it is able to identify related stocks into one sector.
- Our model of Spectral Clustering+KDE outperforms both the S&P 500 and the Markowitz model on the test dataset.
- Pure KDE Model is not able to outperform the Markowitz model on the test dataset. We believe this is because of the curse of dimensionality.
- Intrestingly,the Markowitz model performs better than KDE with and without Spectral Clustering on the training dataset, but not on the test dataset.

Future Directions

- Experiment with different kernels for the KDE, such as Multivariate T distribution
- Experiment with more powerful optimization algorithms such as Adam or SGD with Momentum
- Experiment with improving runtime of the KDE with multiprocessing/GPU acceleration
- Experiment with a different heuristic for choosing the number of sectors
- Find a way to deal with the non-stationarity of the data