Submission Number: 3

Group Number: 3

Group Members:

Full Legal Name	Location (Country)	E-Mail Address	Non-Contributing Member (X)
Shubham Shyamal Kotal	India	shubhkotal@gmail.com	
Ishaan Narula	India	ishaan.narula@outlook.com	

Statement of integrity: By typing the names of all group members in the text box below, you confirm that the assignment submitted is original work produced by the group (*excluding any non-contributing members identified with an "X" above*).

Shubham Shyamal Kotal, Ishaan Narula

Use the box below to explain any attempts to reach out to a non-contributing member. Type (N/A) if all members contributed.

N/A

** Note, you may be required to provide proof of your outreach to non-contributing members upon request.*

Part-3: Discussion – Impacts on EF Under Constraints and Capital Requirements, Fiduciary Responsibilities

Impacts on EF Under Constraints

Note: This has been covered in detail in the Report to Portfolio Manager (Part-3) and the Technical Report on Models Employed (Part-4). Please refer to these reports.

Capital Requirements

A wide range of institutional clients have different capital requirements, based on which their approaches to portfolio management, preferences for different asset classes and constraints vary. We discuss a few of these considerations for insurance companies, pension funds, endowments and foundations and sovereign wealth funds.

Insurance Companies

Insurance companies focus on choosing assets to match the projected probabilistic cash flows of the risks they are underwriting. Accordingly, fixed income assets are typically the largest component of their asset base. In some regions, a book value approach may be the relevant accounting treatment. This renders market price fluctuations irrelevant unless it causes the asset's book value to be written down as OTTI (Other Than Temporary Impairment).

Besides, risk considerations also need to be made to provision for capital for various uses. Considerations include risk-based capital measures, yield, liquidity, potential for forced liquidation of assets to fund claims, credit ratings etc. Regulators normally set minimum capital requirements based on insurers' assets, liabilities and risk, and constrain allocation to certain assets. Many are now moving to Solvency II standards to harmonise regulations across countries.

Pension Funds

Pension funds are subject to a wide array of funding, accounting, reporting and tax constraints, which influence the sponsor's appetite for risk and available asset allocation alternatives. For instance, some countries regulate max or min % in certain asset classes.

Consequently, the sponsor may choose an allocation taking into account funding and financial statement considerations such as anticipated contributions, their volatility, forecasted pension expense/ income under a given allocation scenario etc.

Endowments and Foundations

Endowments and foundations are established with the assumption that they will exist in perpetuity. So, they have a long-term investment horizon. They usually have more flexibility relative to pension funds and insurance companies and can afford higher-risk asset allocations. However, the following two externally-imposed constraints may influence their asset allocation decision:

- Tax Incentives: Many countries provide tax benefits tied to minimum spending requirements (eg, 5% of assets' market value for charitable expenditures otherwise loss of tax-favoured status in the US). These spending requirements may be relaxed if certain types of socially responsible investments are made, which may introduce bias in asset allocation.

- Credit Considerations: Many institutions use endowment or foundation assets to support their balance sheet and borrowing capabilities. Since lenders may require borrowers to maintain certain balance sheet and credit ratios, asset allocation will need to factor in these considerations.

Sovereign Wealth Funds

SWFs are government-owned pools of capital invested on behalf of the peoples of their states with a long-term orientation. They are not generally seeking to defease a set of known liabilities as is common for pension funds and to some extent endowments.

Governing entities of SWFs may adopt regulations constraining the opportunity set for asset allocation. Besides, SWFs invest more conservatively than appropriate given their long-term investment horizon, since they are subject to public scrutiny and would like to reduce reputation risk.

Cultural and religious factors also constrain asset allocation choices. For instance, Sharia law prohibits investment in businesses whose core activities are considered vices in Islam (pork, alcohol etc.).

Lastly, ESG considerations are also becoming increasingly important in asset allocation decisions. However, these are not modelled during the decision process. Consequently, they may be achieved through the implementation of asset allocation or through an amount of capital set aside for these purposes.

Fiduciary Responsibilities

A portfolio manager has a fiduciary responsibility to be loyal to their clients, act in their best interest, and exercise reasonable care and prudent judgment.

Per the CFA Institute, fiduciary duties are often imposed by law or regulation when an individual or institution is charged with the duty of acting for the benefit of another party, such as managing investment assets. The duty required in fiduciary relationships exceeds what is acceptable in many other business relationships because a fiduciary is in an enhanced position of trust.

To fulfil their fiduciary duties, the first step for a portfolio manager or an investment adviser is to identify who their client is, to whom a duty of loyalty is owed. In the absence of a client or a beneficiary (e.g. a PM managing a fund to an index), the duty of loyalty, prudence and care must be carried out in a manner consistent with the stated mandate.

The next step is portfolio construction and management, whereby the PM must make all efforts to gauge the level of knowledge of the client, the goals they are looking to achieve and their risk tolerance. Only after a thorough understanding of these areas should the PM devise an appropriate portfolio, based on capital market expectations and client constraints. The PM must also make every effort to educate the client in the process.

Lastly, soft commission and proxy voting policies are other areas where the PM is expected to fulfil their fiduciary duty of loyalty, prudence and care. Conflicts may arise when a PM uses client brokerage to purchase research services. This practice is referred to as soft commissions, and a PM who pays a higher commission than what they would normally pay for such services, without corresponding benefit to the client, is violating their fiduciary duty. Finally, a PM who fails to vote proxies with sufficient consideration and proper judgement also violates their fiduciary obligations to their client, since proxies have economic value to clients.

Part-3: Discussion – Report to the Portfolio Manager

Note: All of the findings in this part have also been covered in the technical report in part-4

Introduction

We wrap up our ETF research by examining the 165 portfolio combinations we created earlier, drawing three sample 3-ETF portfolios, and comparing the portfolio's efficient frontier to an efficient frontier under long-only constraints and an efficient frontier under box constraints of 10% minimum weight and 40% maximum weight. The nine efficient frontiers are depicted here, along with the portfolio combination that was used to create the frontier.

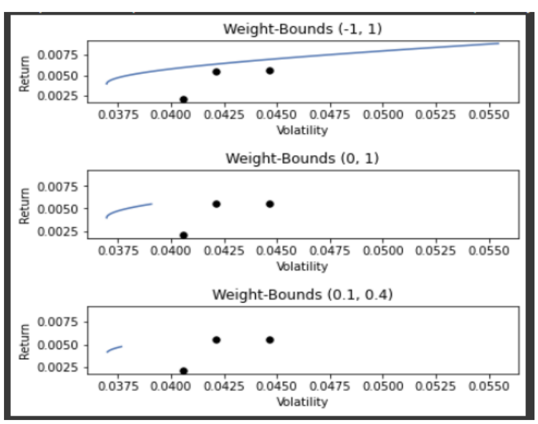
Efficient Frontier Under Constraints

In the absence of any constraints on expected return or variance of the portfolio or on weights of the various asset classes, optimal portfolios can be determined using an analytical closed-form formula. Although mathematically elegant, the resulting portfolio may often require heavy use of leverage to finance its holdings and may be of little use in the real world to several investors. This analytic calculation of the optimal portfolio weights is not possible in the presence of constraints, as a result of which one has to resort to numerical methods for derivation of the efficient frontier. However, introducing such constraints results in a frontier which is customised to the special needs of the investor and, therefore, provides much more utility.

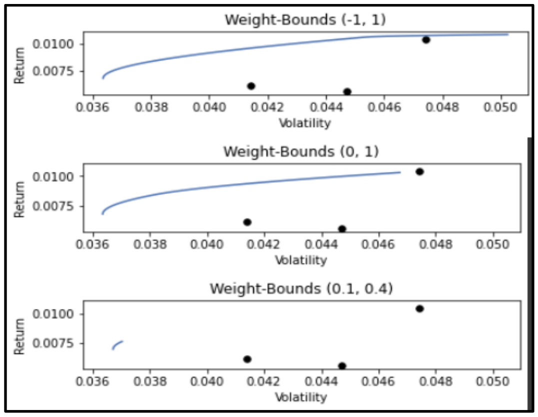
Typically, constraining leverage levels shifts the efficient frontier inwards and makes it a finite curve. An interesting point to note is that throughout most of the intermediate range of returns, the decline in return resulting from constraining leverage levels, for a given level of risk, is quite minimal when compared with the large difference between unlimited leverage and constrained leverage. Introducing constraints on the position sizes of various assets narrows down the range of returns available on the efficient frontier. Nevertheless, optimal portfolios on this frontier are usually not far from those on the original unconstrained frontier for the narrowed return range.

Key Project Findings

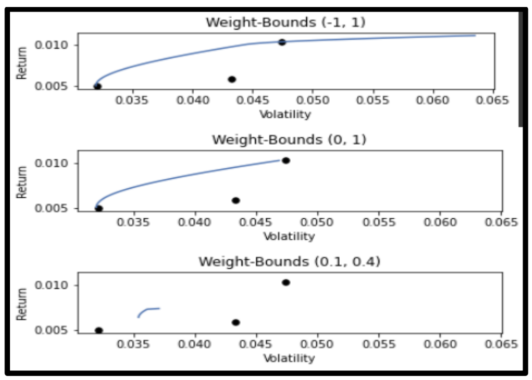
IYR, IYZ, XLB



XLB, XLK, XLU



XLI, XLK, XLP



The EF with no constraints provides the widest range of returns and volatility. Because negative holding permits for short holdings, this is the case. Volatility rises in tandem with greater rewards. The capital required for a no-constraint portfolio should be substantially larger than for a long-only portfolio due to the leverage given by short holdings. The increasing cost of borrowing capital should be considered into the portfolio construction of the asset manager. This indicates that instead of a constant risk-free interest rate, a higher borrowing rate may be necessary in the optimization.

The long-only EF is roughly the left part of the no-constraint EF because the EF return cannot exceed the largest return of the underlying three ETFs. For any goal volatility, the best feasible return for long-only EF is somewhat lower than the highest potential return for no-constraint EF (x-axis).

The min-max (0.1-0.4) EF has a substantially narrower range of return volatility due to the set range of weights. The use of weight limitations could lead to a more stable EF portfolio. This could be a beneficial feature given the ambiguity around the input assumptions (expected return and covariance). Weight bounds may be required due to fiduciary responsibility, which requires the asset manager to advise the client on investment strategy while taking into consideration any investment constraints the client may have.

Conclusion

Therefore, it helps to incorporate as many practical constraints as possible on the underlying portfolio model in order to obtain portfolios that are meaningful to the investor.

Part-3: Discussion – Work Split Report

Part-1: Data Importing	Shubham
Part-2: Category 1 Constrained Portfolios	Shubham
Part-3: Discussion	Ishaan
Part-4: Technical Report on Models Employed	
- Introduction	Ishaan
Submission-1 Models	
- Linear Regression	Shubham
- Lasso Regression	Shubham
- Regression Tree	Shubham
- K-Means Clustering	Ishaan
Submission-2 Models	
- Principal Component Analysis	Ishaan
- Markowitz Model and the Efficient Frontier	Ishaan
- Robust Techniques for Efficient Frontier Estimation – Covariance Shrinkage and the Critical Line Algorithm	Ishaan
Submission 3 Models	
- Efficient Frontier Estimation under Constraints	Ishaan
- Conclusion	Shubham & Ishaan
- References	Shubham & Ishaan

Part-4: Technical Report on Models Employed

Outline

- Introduction

Submission-1 Models

- Linear Regression
- Lasso Regression
- Regression Tree
- K-Means Clustering

Submission-2 Models

- Principal Component Analysis
- Markowitz Model and the Efficient Frontier
- Robust Techniques for Efficient Frontier Estimation – Covariance Shrinkage and the Critical Line Algorithm

Submission 3 Models

- Efficient Frontier Estimation under Constraints

- Conclusion
- References

MScFE 640 Portfolio Theory and Asset Pricing

Group #: 3 Submission #: 3

Introduction

Portfolio management is one of the central areas of research in modern finance. In this report, we cover the technical details underlying the statistical and portfolio optimisation models as well as how these were applied in the group project within this course. We advise to read this report after or in conjunction with the Jupyter notebooks containing our analyses.

Our group project uses data on 11 ETFs covering various sectors of the US economy as the primary dataset. We began the project by looking at the 4-year monthly return history over 2014-2019 of these ETFs and trying to find relationships between these and leading, coincident and lagging economic indicators. We employed towards the **linear and lasso regression models, decision trees and K-means clustering algorithm** towards this end.

Thereafter, we focused our attention to 2019 and 2020 daily returns and analysed portfolio statistics and efficient frontiers for 2-asset (XLK, XLI) and 3-asset (XLK, XLI and XLY) portfolios of our choice. We then extended this analysis to all 3-ETF portfolios which could be formed from our 11-ETF dataset. Thereafter, the economic indicator (LEI, CEI, LAG) and positional (long + / short -) categories for each ETF in each portfolio were specified, to look for indicator buckets which tended to be positively associated with returns. Finally, we constructed a new portfolio from the first 3 principal components of our 11-dimensional dataset and compared its return performance over 2019 and 2020 with our 3-ETF portfolios. Throughout this analysis, we employed the **Principal Component Analysis, Markowitz model, covariance shrinkage and the Critical Line Algorithm**.

In the third submission, we analysed and compared the **efficient frontiers for all portfolio combinations under no constraints, long-only constraints, and position size constraints**.

Linear Regression

To determine the strength of predictors, direction, and predict an event, a linear regression is utilized. A positive correlation means that when one variable travels in one direction, the other variable moves in the same direction. This direction might be positive or negative, but both variables must move in the same direction. The variables move in opposite directions when there is a negative correlation. A regression coefficient goes from -1 to 1, with the stronger the association the closer it comes to 1 or -1. A correlation of 1 or -1 indicates a 100 percent correlation, meaning both variables will move by the same percentage. A correlation of 0 shows that there is no relationship between the variables. We're looking for correlations between the movement of the economic indicators and the ETFs we've chosen.

There are four assumptions associated with a linear regression model:

- Linearity: The relationship between X and the mean of Y is linear
- Homoscedasticity: The variance of residual is the same for any value of X
- Independence: Observations are independent of each other
- Normality: For any fixed value of X, Y is normally distributed

To find the daily returns, we first calculated the percent change in each of the selected SPDR funds. We can observe how the funds performed from January 2014 to December 2019 by looking at the daily returns. We next compared each fund's daily return and examined them four times. Once against all 21 economic indicators, once against the 10 leading indicators, once against the 7 lagging indicators, once against the 4 coincidental indicators, and once against the 10 leading indicators. The outcomes of each regression are listed below.

	key	leading	lagging	coincidental	all_factors
0	XLF	0.103877	0.038322	-0.027935	0.278845
1	XLU	-0.114475	0.032838	-0.043479	-0.055258
2	XLRE	-0.143189	-0.092884	-0.037506	-0.082523
3	XLE	0.010324	-0.025481	0.003880	0.195588
4	XLB	-0.016412	-0.060567	-0.032117	0.140692
5	XLI	0.032350	-0.037099	-0.044180	0.138938
6	XLK	0.063791	-0.054441	-0.047886	0.223882
7	XTL	-0.023704	-0.031847	-0.035647	0.041508
8	XLP	0.023632	-0.041952	-0.036785	0.092095
9	XLY	0.002995	-0.076697	-0.043873	0.211159
10	XLV	0.104133	-0.050264	-0.047336	0.244279

Analyzing the Models

We can conclude that the highest correlation arises when we study the fund's movement using all 21 economic variables by comparing the results in each row. Aside from XLU and XLRE, which had the best performance utilizing all 21 economic indicators, nine other ETFs had the best results using only the lagging and coincidental indicators. This is most likely due to the fact that including all variables results in a more balanced regression because it represents the overall economy affecting the ETF.

Because the financial world is complex and there are several elements that push and impact the market, employing only a portion of the economic factors will not provide us with a comprehensive picture of the economy, resulting in a poorer analysis. Apart from using all of the indicators, we can check which ones forecasted and displayed the best model for each ETF in the table below.

For the Financial, Industrial, Technology, Consumer Staples, Consumer Discretionary, and Health Care Select Sector SPDR Fund, the LEI economic indicators were the best model. For the Real Estate, Energy, and Materials sectors, the CEI economic indicators were the best model, while the LAG indicators were the best for Utilities and Telecom Fund.

	key	leading	lagging	coincidental
0	XLF	0.103877	0.038322	-0.027935
1	XLU	-0.114475	0.032838	-0.043479
2	XLRE	-0.143189	-0.092884	-0.037506
3	XLE	0.010324	-0.025481	0.003880
4	XLB	-0.016412	-0.060567	-0.032117
5	XLI	0.032350	-0.037099	-0.044180
6	XLK	0.063791	-0.054441	-0.047886
7	XTL	-0.023704	-0.031847	-0.035647
8	XLP	0.023632	-0.041952	-0.036785
9	XLY	0.002995	-0.076697	-0.043873
10	XLV	0.104133	-0.050264	-0.047336

Comparing Regressions

Linear regression is the most fundamental and widely used method for determining the strength of a correlation between variables. The formula for linear regression is usually written as follows.

- Simple linear regression: Y = a + bX + u
- Multiple linear regression: Y = a + b1X1 + b2X2 + b3X3 + ... + btXt + u

  where

  Y = the variable that you are trying to predict (dependent variable)

  X = the variable that you are using to predict Y (independent variable)

  a = the intercept

  b = the slope

  u = the regression residual

Linear regression implies that the variables have a pre-existing linear relationship and attempts to uncover it. The b variable is distinct for each X variable, as you can see above. This implies we may give each variable varying weights, and if one variable has a bigger influence on the predictor variable, we can give it a higher weight for a more accurate assessment. If we have a tiny dataset, this can backfire and lead to overfitting in some circumstances. Lasso regression is a linear regression derivative, and its equation is shown below.

$$\frac{1}{2m}\sum_{i=1}^{m}(y-Xw)^2 + alpha\sum_{j=1}^{p}\left|w_j\right|$$

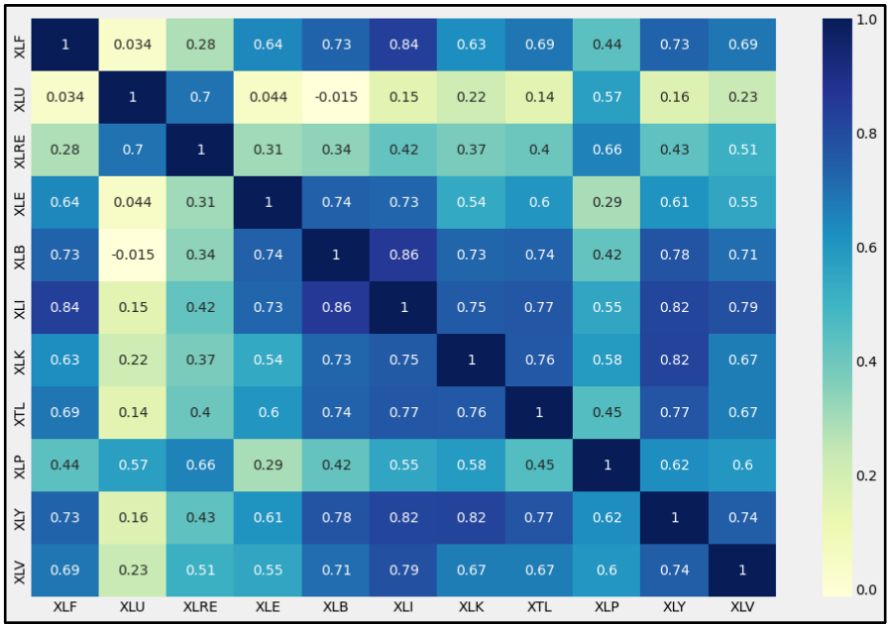
Lasso Regression

Lasso is a variant of linear regression in which the model is penalized for the sum of the weights' absolute values. As a result, the absolute values of weight will be reduced (in general), and many will be zeros. The alpha in the equation above helps to mitigate the effects of having too many weights on each variable. The Lasso regression attempts to normalize the equation by reducing the higher weights to a lower amount; however, because the alpha affects all weights equally, the already small weights will be decreased to zero, thus removing them. When we compare the lasso regression results to the linear regression findings, we can observe that the linear regression clearly outperforms the lasso regression. The lasso regression values were all negative and lower than the regression figures based on all economic indices.

	leading	lagging	coincidental	all_factors	total	all_factors_lasso
0	0.103877	0.038322	-0.027935	0.278845	0.114264	-0.267998
1	-0.114475	0.032838	-0.043479	-0.055258	-0.125117	-0.401383
2	-0.143189	-0.092884	-0.037506	-0.082523	-0.273579	-0.702546
3	0.010324	-0.025481	0.003880	0.195588	-0.011277	-0.303514
4	-0.016412	-0.060567	-0.032117	0.140692	-0.109096	-0.359340
5	0.032350	-0.037099	-0.044180	0.138938	-0.048928	-0.342748
6	0.063791	-0.054441	-0.047886	0.223882	-0.038536	-0.268153
7	-0.023704	-0.031847	-0.035647	0.041508	-0.091197	-0.397537
8	0.023632	-0.041952	-0.036785	0.092095	-0.055105	-0.388177
9	0.002995	-0.076697	-0.043873	0.211159	-0.117575	-0.363510
10	0.104133	-0.050264	-0.047336	0.244279	0.006533	-0.233841

Regression Tree

We can see that Linear regression outperformed Lasso regression, but we must also examine the correlation between the ETFs, as large correlations will alter how each ETF affects the others. To see how each ETF interacts with each other, we compute the correlation matrix and its corresponding heat map.



The XLU and XLRE rows have the most of the low correlation and darker areas on the heat map. When XLU and XLRE are removed from the equation, the remaining ETFs have a correlation of at least 50%, and the graph would be primarily lighter colors, indicating a correlation closer to 1. This means that as the majority of the ETFs move, whether positively or negatively, the other sectors will be affected and will respond in kind. Finally, we use decision trees to compare our ETFs, running each one through a tree that includes all 21 economic indices. The results are shown below.

	leading	lagging	coincidental	all_factors	total	trees
0	0.103877	0.038322	-0.027935	0.278845	0.114264	1.000000
1	-0.114475	0.032838	-0.043479	-0.055258	-0.125117	1.000000
2	-0.143189	-0.092884	-0.037506	-0.082523	-0.273579	1.000000
3	0.010324	-0.025481	0.003880	0.195588	-0.011277	1.000000
4	-0.016412	-0.060567	-0.032117	0.140692	-0.109096	1.000000
5	0.032350	-0.037099	-0.044180	0.138938	-0.048928	1.000000
6	0.063791	-0.054441	-0.047886	0.223882	-0.038536	1.000000
7	-0.023704	-0.031847	-0.035647	0.041508	-0.091197	1.000000
8	0.023632	-0.041952	-0.036785	0.092095	-0.055105	1.000000
9	0.002995	-0.076697	-0.043873	0.211159	-0.117575	1.000000
10	0.104133	-0.050264	-0.047336	0.244279	0.006533	1.000000

This time, the outcome is different, as the decision tree method looks to produce superior results. The drawback is that it appears to show a severe case of overfitting. As a result, it must be used in conjunction with other elements in practice. Overall, we can see that linear regression was the most effective method for determining the strength of the correlation between variables. Although the regression trees predicted a value of 1 for each ETF, the results are untrustworthy, and more research is needed to properly anticipate outcomes.

K-Means Clustering

Distance correlation is a robust and universal way to check if there is a relation, linear or non-linear, between two numeric variables. For example, if we have a set of pairs of numbers: (x1, y1) (x2, y2) ... (xn, yn), we can use distance correlation to check if there is any (not necessarily linear) relation between the two variables (x and y). Moreover, x and y can be vectors of different dimensions.

To calculate distance correlation, we first use xi to calculate distance matrix. Then we calculate distance matrix using yi. The two distance matrices will have the same dimensions because the number of xi and yi is the same (because they come in pairs). This gives us a set of pairs of distances and we can use it to calculate distance correlation (correlation between distances).

	iyr log return	iyz log return	xlb log return	xle log return	xlf log return	xli log return	xlk log return	xlp log return	xlu log return	xlv log return	xly log return
iyr log return	1.0	0.364468	0.327421	0.27978	0.299049	0.366434	0.375585	0.544865	0.629949	0.501191	0.434098
iyz log return	0.364468	1.0	0.614839	0.492231	0.472404	0.546916	0.548503	0.412458	0.260671	0.499889	0.57113
xlb log return	0.327421	0.614839	1.0	0.685077	0.697103	0.823083	0.675631	0.405641	0.162361	0.596323	0.73264
xle log return	0.27978	0.492231	0.685077	1.0	0.549166	0.595779	0.452804	0.301993	0.153785	0.359273	0.537837
xlf log return	0.299049	0.472404	0.697103	0.549166	1.0	0.774958	0.611338	0.431736	0.170647	0.605111	0.672079
xli log return	0.366434	0.546916	0.823083	0.595779	0.774958	1.0	0.681413	0.539493	0.179869	0.655141	0.746988
xlk log return	0.375585	0.548503	0.675631	0.452804	0.611338	0.681413	1.0	0.493577	0.242756	0.539397	0.748096
xlp log return	0.544865	0.412458	0.405641	0.301993	0.431736	0.539493	0.493577	1.0	0.533488	0.478115	0.577819
xlu log return	0.629949	0.260671	0.162361	0.153785	0.170647	0.179869	0.242756	0.533488	1.0	0.212626	0.212824
xlv log return	0.501191	0.499889	0.596323	0.359273	0.605111	0.655141	0.539397	0.478115	0.212626	1.0	0.657767
xly log return	0.434098	0.57113	0.73264	0.537837	0.672079	0.746988	0.748096	0.577819	0.212824	0.657767	1.0

The above distance correlation matrix, calculated based on returns of the 11 ETFs, shows that there is a strong relationship between a majority of the ETFs. Because of this, one can imagine that returns across the 11 ETFs are governed by a common set of underlying factors.

However, the linear model showed that the 19 economic indicators combined explain only a limited part of the total variability (13-30%) in the ETFs' returns.

The above findings could imply the following:

- either the selected economic indicators are not a good choice of predictors to explain the ETFs' returns or
- if they are, the true underlying relationship between the ETFs' returns and the economic indicators is non-linear

We then ran a K-means clustering with K=3 and find that 9 of the ETFs are placed in one cluster whereas XLE (Energy) and XLV (Healthcare) are in two separate clusters.

Principal Component Analysis

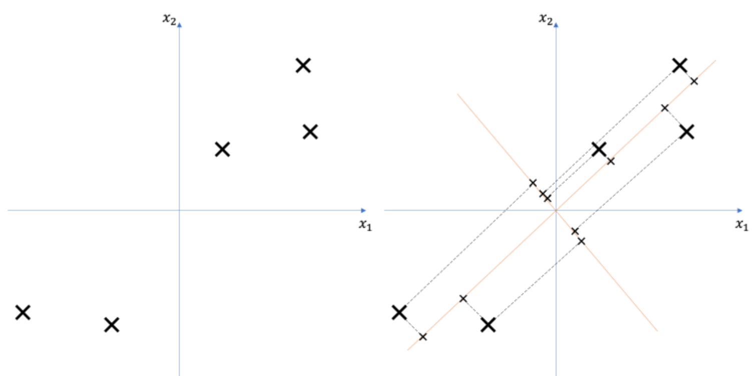
Overview

Principal Component Analysis (PCA) is an unsupervised learning technique for dimensionality reduction. The idea of dimensionality reduction is to find a way to transform an unlabelled dataset with, say *n* features or dimensions into a dataset with a fewer number of dimensions, say *k*, while trying to minimise the information loss during the compression process.

The Model

Problem Formulation

To facilitate intuitive understanding, we formulate the problem which PCA solves with 2 features before progressing to the more general *n*-dimensional case. Suppose we have a 2-dimensional dataset as shown in the chart below, and we would like to reduce all the information it contains into to a 1-dimensional dataset.



So, the idea is essentially to find a straight line onto which each ordered pair can be projected in such a way that the resulting projected point captures all the information contained in that ordered pair. Mathematically, this is achieved through PCA when the ordered pair is projected perpendicularly onto the straight line, and the straight line is such that this perpendicular distance between the ordered pair and the projected point is minimised. Stated alternatively, PCA tries to find a lower dimensional linear subspace, a line in this case, onto which to project the data such that the sum of squares of orthogonal projection errors is minimised.

That said, the upward sloping line drawn through the points may be a good choice for the lower dimensional surface onto which to project the data, when compared to the downward sloping line, since the orthogonal projection errors on the former are smaller compared to those on the latter.

More generally, to reduce an *n*-dimensional dataset into a *k*-dimensional dataset, PCA involves finding a *k*-dimensional linear subspace (which essentially means finding *k* vectors which span this subspace) onto which to project the *n*-dimensional datapoints such that the sum of squares of orthogonal projection errors is minimised.

The Algorithm

It follows from the above discussion that we need to compute two quantities to implement PCA:

- the direction of the vectors which span the lower-dimensional linear subspace onto which the higher-dimensional datapoints are projected. These are given by eigenvectors of the covariance matrix
- the projections of the datapoints onto the lower-dimensional linear subspace

The following is an algorithm of the PCA technique. We assume we have *m* datapoints each of which is *n*-dimensional.

Step-1: Before implementing PCA, it is standard practice to carry out mean normalisation and feature scaling, especially when the *n* features have vastly different means and scales.

Mean normalisation entails computing the mean of the data recorded for a given feature and subtracting that from the data. As a result, the data on each feature would have approximately zero mean.

Feature scaling involves dividing each mean-normalised data point for a given feature by a measure of dispersion (range, standard deviation etc.) in that feature's data

Step-2: Compute the covariance matrix. For each training example, this basically entails multiplying every feature with itself and with other features and storing the products into an (1 x *n*) row vector. This way, we have *n* such row vectors, one for each feature.

We repeat the above for every training example and take the average across these *m* matrices.

$$\Sigma = \frac{1}{m-1}\sum_{i=1}^{m}\left(x^{(i)}\right).\left(x^{(i)}\right)^{T}$$

Step-3: Compute the eigenvectors of the covariance matrix. These can be computed using Singular Value Decomposition (SVD)/ diagonalisation of the covariance matrix. SVD is a matrix factorisation method which extends diagonalisation, i.e. use of eigenvalues and eigenvectors to factor a matrix into a more useful product of matrices, to non-diagonalisable matrices. Since our covariance matrix $\sum$ is symmetric and therefore diagonalisable, it can be factorised as follows:

$$\sum = VLV^{T}$$

where
V: matrix of eigenvectors. These are the principal axes or the principal directions of the data
L: diagonal matrix with eigenvalues λ_i in decreasing order on the diagonal

Note: The above SVD procedure can also be directly carried out on the mean-normalised data matrix of size *m x n* to get the matrix of principal directions. Performing SVD on the data matrix X gives us

$$X = USV^{T}$$

where
U: unitary matrix
S: diagonal matrix of singular values s_i
V: matrix of eigenvectors. These are the principal axes or the principal directions of the data

We now relate the results of the SVD procedure on the covariance matrix with that on the data matrix, to show that they lead to the same result.

Since the covariance matrix, in vectorised form, is given by

$$\Sigma = \frac{1}{(m-1)}X^{T}X$$

, it follows that

$$\Sigma = \frac{1}{(m-1)} VSU^T USV^T$$

$$\Sigma = V \frac{S^2}{(m-1)} V^T$$

Therefore, SVD of the data matrix also leads to the conclusion that V provides the principal axes or the principal directions of the data. The singular values are related to the eigenvalues of the covariance matrix by

$$\lambda_i = \frac{{s_i}^2}{(m-1)}$$

Step-4: To get the *k* dimensions, we select the first *k* columns of the V matrix which yields the *n x k* matrix V_{reduce}. Finally, to create a data point in *k* dimensions, we multiply the transpose of V_{reduce} (size *k x n*) by that data point in *n* dimensions, and the result is a *(k x n) x (n x 1) --> (k x 1)* vector, which is the reduced data point.

$$z = {V_{reduce}}^T x$$

In vectorised form, the *k* principal components are given by

$$XV = USV^T V = US$$

Step-5 *(Optional)*: One can also go back from a compressed representation of the data to an approximation of the original high-dimensional data (i.e. reconstruction). As shown previously, we obtain the compressed vector *z* using the below formula

$$z = {V_{reduce}}^T x$$

In order to obtain an approximation of the original vector *x*, we simply multiply both sides of the above equation by the transpose of V_{reduce} as follows

$$V_{reduce}\, z = V_{reduce}.{V_{reduce}}^T x_{approx}$$

$$x_{approx} = V_{reduce}\, z$$

Graphically, the resulting vector x_{approx} is a point on the linear subspace which minimises the projection error (perpendicular distance).

General Use Cases

One may want to undertake dimensionality reduction for two reasons:

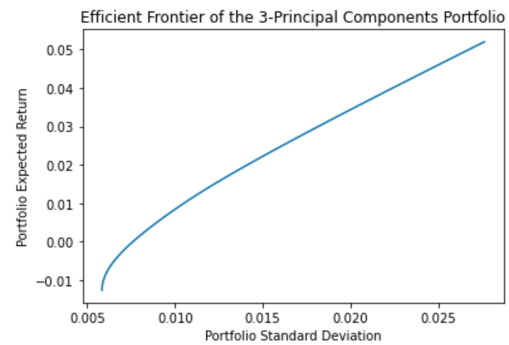
- Data Compression: which may enable one to eliminate redundant features in the dataset, reduce the amount of storage space which the compressed dataset takes in computer memory and, thereby, speed up learning algorithms

- Data Visualisation: One may be interested in visually representing a high-dimensional dataset, for which dimensionality reduction may be required since it is only possible to visualise data in at most 3 dimensions at a time

Application in Project

We use daily return data on the 11 ETFs covering various sectors of the US economy over 2019 as our dataset. This comprises of 251 data points, each of which is 11-dimensional. This dataset is reduced to a 3-dimensional dataset using Principal Component Analysis.

The 3 Principal Components are then used to estimate 2019 expected return (taken to be the CAGR over 2019) and covariances, which are then fed into the Critical Line Algorithm (CLA) to obtain the efficient frontier (chart alongside) resulting from a portfolio of the 3 PCs.



We then use the constant volatility level of 0.8% to get the expected return and weights for the 3-PC portfolio lying on the efficient frontier for this assumed volatility level. The return of this PCA portfolio is 0.176% p.a. which is much lower than the average of the 2019 returns across 165 combinations of 3-asset portfolios, which stands at 23.566%.

Thereafter, PCA is applied on 2020 daily return data for the 11 ETFs to get 3 Principal Components, and historical estimates of expected returns are calculated for each of these 3 PCs. Then, these expected returns are used with the weights of the 2019 PCA portfolio which were obtained from the efficient frontier for the assumed volatility of 0.8%, to get the expected return for the 2020 PCA portfolio. This stands at 3.89%, higher than the expected return for the 2019 PCA portfolio.

The Efficient Frontier and the Markowitz Model

Overview

Portfolio optimisation is one of the cornerstones of modern quantitative finance. At its foundation lies the Markowitz model which propounds portfolios with the most favourable risk-return trade-off by optimisation of portfolio expected returns and standard deviation (Mean-Variance optimisation). Although the model suffers from many limitations, it is a great starting point for more realistic models. We first derive the efficient frontier and depict it diagrammatically, following which the Markowitz Model is discussed.

Model

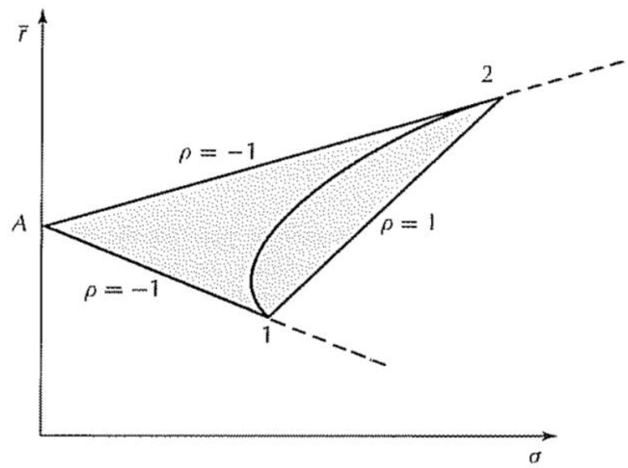
Portfolio Diagram with 2 Assets

To derive the efficient frontier, we start by visualising how a portfolio of assets can be represented on a 2-dimensional diagram with mean return on one axis and standard deviation of returns on the other.

For the sake of simplicity, consider 2 assets whose returns have mean and standard deviation represented by points 1 and 2 in the diagram below. The mean and standard deviation of returns of a portfolio created with weight α assigned to asset 2 and $(1-\alpha)$ assigned to asset 1 can be calculated based on 3 variables: the means and standard deviations of the individual assets' returns, and the correlation between those returns. This is shown below:

$$\bar{r}(\alpha) = (1-\alpha)\bar{r}_1 + \alpha\bar{r}_2$$

$$\sigma(\alpha,\rho) = \sqrt{(1-\alpha)^2\sigma_1^2 + 2\rho\alpha(1-\alpha)\sigma_1\sigma_2 + \alpha^2\sigma_2^2}$$



Given the mean returns for assets 1 and 2, the portfolio's return depends only on α whereas its standard deviation depends on α and ρ.

More specifically, the portfolio's return is a linear combination of the returns of assets 1 and 2 (i.e. in a portfolio made up with a 50-50 mix, for example, the portfolio's return would lie midway between the individual assets' returns). The portfolio return's standard deviation, however, is a linear combination of those of assets 1 and 2 only when $\rho = 1$ or $\rho = -1$. We look at how the portfolio return's standard deviation varies with α under these 2 cases for ρ:

<u>Case-1: $\rho = 1$</u>

$$\sigma(\alpha, 1) = \sqrt{(1-\alpha)^2\sigma_1^2 + 2\alpha(1-\alpha)\sigma_1\sigma_2 + \alpha^2\sigma_2^2}$$

$$\sigma(\alpha, 1) = \sqrt{[(1-\alpha)\sigma_1 + \alpha\sigma_2]^2}$$

$$\sigma(\alpha, 1) = (1-\alpha)\sigma_1 + \alpha\sigma_2$$

Diagrammatically, when $\rho = 1$, both the portfolio return's mean and standard deviation are linear combinations of those for assets 1 and 2. So, in this case, the portfolio traces the direct straight line shown on the diagram above as α varies.

<u>Case-2: $\rho = -1$</u>

$$\sigma(\alpha, 1) = \sqrt{(1-\alpha)^2\sigma_1^2 - 2\alpha(1-\alpha)\sigma_1\sigma_2 + \alpha^2\sigma_2^2}$$

$$\sigma(\alpha, 1) = \sqrt{[(1-\alpha)\sigma_1 - \alpha\sigma_2]^2}$$

$$\sigma(\alpha, 1) = \begin{cases} (1-\alpha)\sigma_1 - \alpha\sigma_2 \; if \; \alpha < \dfrac{\sigma_1}{\sigma_1 + \sigma_2} \\ -[(1-\alpha)\sigma_1 - \alpha\sigma_2] \; if \; \alpha > \dfrac{\sigma_1}{\sigma_1 + \sigma_2} \end{cases}$$

$$\sigma(\alpha, 1) = |(1-\alpha)\sigma_1 - \alpha\sigma_2|$$

When $\rho = -1$, the portfolio return's mean is a linear combination of those for assets 1 and 2. The portfolio return's standard deviation is now a linear combination of:

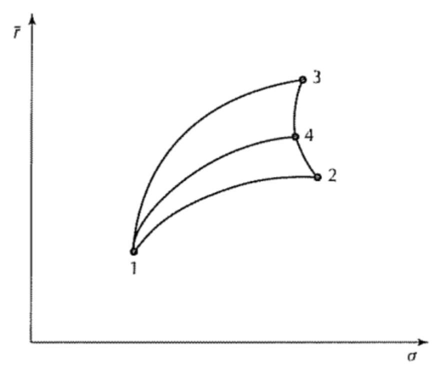
- The standard deviation of asset 1's return and negative of the standard deviation of asset 2's return when $\alpha < \frac{\sigma_1}{\sigma_1+\sigma_2}$. So, the portfolio traces the straight line between point 1 and the mirror image of point 2 in the diagram.
- Negative of the standard deviation of asset 1's return and the standard deviation of asset 2's return when $\alpha > \frac{\sigma_1}{\sigma_1+\sigma_2}$. So, the portfolio traces the straight line between the mirror image of point 1 and point 2 in the diagram.

Since the portfolio traces out a straight line between assets 1 and 2 under one extreme of $\rho = 1$ and a kinked curve under the other extreme of $\rho = -1$, we conclude that it would trace out a smooth curve connecting points 1 and 2 for an intermediate value of ρ, as α varies from 0 to 1. When the weights are negative (either asset 1 or asset 2 is shorted), the portfolio would lie somewhere on the dashed lines shown in the diagram.

Feasible Set/ Region

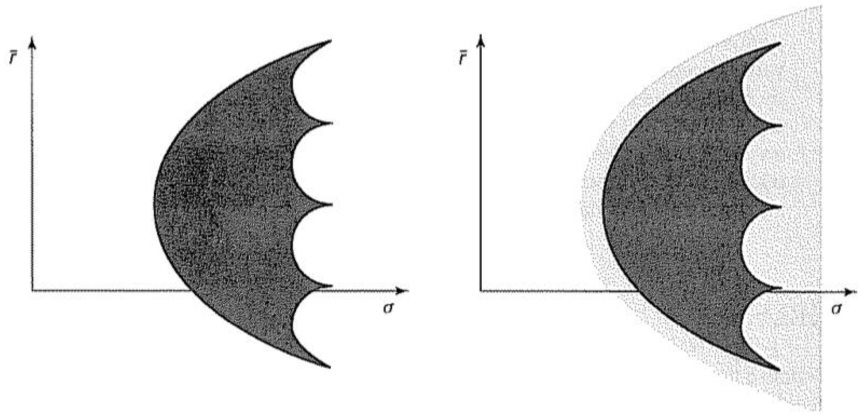
Extending the above analysis to *n* assets, the portfolios formed all possible combinations of these *n* assets using every possible weighting scheme would comprise the feasible set or the feasible region. This set has 2 properties:

Property-1: If there are at least 3 imperfectly correlated assets with different means and standard deviations, the feasible set would be a solid 2-dimensional region.



As shown in the above diagram of 3 assets, this is true because 2-asset portfolios of assets 1 and 2, 2 and 3, and 1 and 3 would be found on the 3 curves connecting those points respectively. Thereafter, one way of forming a 3-asset portfolio could be to take a portfolio of assets 2 and 3, in this case point 4 on the diagram, combine it with asset 1 (i.e. curve connecting points 1 and 4) and then allow the weights of assets 2 and 3 in portfolio 4 to vary so that the curve connecting points 1 and 4 traces out the solid region enclosed by the curves connecting points 1, 2 and 3.

Property-2: The feasible region is convex to the left, meaning that a straight line connecting any two points in the feasible region would never cross the region's left boundary. This is shown in the diagram below.

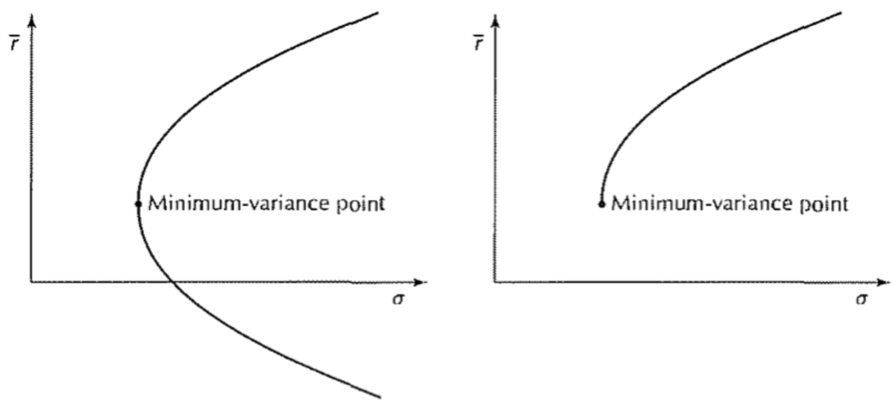


The above is true because all portfolios made from two assets (with positive weights) lie on or on the left of the line connecting them. Furthermore, the feasible region with short selling encompasses the feasible region without short selling.

Minimum-variance Set and Efficient Frontier

The minimum-variance set and the efficient frontier are parts of the feasible region. We derive these making two behavioural assumptions about investors:

- Risk Aversion: Given the level of return, they seek portfolios to minimise overall risk (captured by portfolio standard deviation)
- Non-satiation: Given the level of risk, investors seek to achieve the highest possible rate of return



The left boundary of the feasible set is called the minimum-variance set, since for a given mean rate of return, the portfolio on the boundary for that mean return has the lowest standard deviation. Risk-averse investors would prefer these points over others in the feasible region. Out of all points on the minimum-variance set, the portfolio on the point with the lowest standard deviation is called the minimum-variance point (MVP).

Out of all points in the minimum-variance set, the ones on the upper part (i.e. above the MVP) offer the highest return for a given level of risk (standard deviation). This is the efficient frontier of the feasible region, since it represents the set of portfolios which offer the most favourable risk-return trade-off.

The Markowitz Model

The Markowitz model provides a mathematical structure to the problem of finding portfolios lying on the efficient frontier. The solution to the Markowitz model is an analytical representation of the minimum variance portfolio for a given mean rate of return.

Mathematically, in a portfolio of *n* assets, the objective function to be minimised is the variance of that portfolio subject to two constraints: the portfolio has a fixed rate of return and the weights across the individual assets sum to unity. What follows is the problem statement of the Markowitz model.

$$minimize\ \frac{1}{2}\sum_{i,j=1}^{n} w_i w_j \sigma_{ij}$$

$subject\ to$

$$\sum_{i=1}^{n} w_i \bar{r}_i = \bar{r}$$

$$\sum_{i=1}^{n} w_i = 1$$

Conditions for the solution to this problem can be found using Lagrange multipliers λ (for constraint 1) and μ (for constraint 2). The Lagrangian function is as follows:

$$L = \frac{1}{2}\sum_{i,j=1}^{n} w_i w_j \sigma_{ij} - \lambda\left(\sum_{i=1}^{n} w_i \bar{r}_i - \bar{r}\right) - \mu\left(\sum_{i=1}^{n} w_i - 1\right)$$

The above function is differentiated w.r.t each variable w_i and the resulting derivative is set to zero. We present the equations for the 2-asset case.

$$L = \frac{1}{2}(w_1^2\sigma_1^2 + w_1 w_2 \sigma_{12} + w_2 w_1 \sigma_{21} + w_2^2 \sigma_2^2) - \lambda(\bar{r}_1 w_1 + \bar{r}_2 w_2 - \bar{r}) - \mu(w_1 + w_2 - 1)$$

Differentiating w.r.t w_1 and w_2

$$\frac{\partial L}{\partial w_1} = \frac{1}{2}(2\sigma_1^2 w_1 + \sigma_{12} w_2 + \sigma_{21} w_2) - \lambda \bar{r}_1 - \mu$$

$$\frac{\partial L}{\partial w_2} = \frac{1}{2}(\sigma_{12} w_1 + \sigma_{21} w_1 + 2\sigma_2^2 w_2) - \lambda \bar{r}_2 - \mu$$

We now have the below first order conditions:

$$\frac{\partial L}{\partial w_1} = 0 \qquad \frac{\partial L}{\partial w_2} = 0$$

We also know that $\sigma_{12} = \sigma_{21}$. Based on these conditions, we get the following linear equations:

$$\sigma_1^2 w_1 + \sigma_{12} w_2 - \lambda \bar{r}_1 - \mu = 0$$

$$\sigma_{21} w_1 + \sigma_2^2 w_2 - \lambda \bar{r}_2 - \mu = 0$$

Using the above two equations with the two equality constraints, we get a system of 4 linear equations with 4 unknowns which can be solved using standard linear algebra methods. Similarly, for the *n*-assets case, we would have a system of *n+2* linear equations with *n+2* unknowns, as shown below.

$$\sum_{j=1}^{n} \sigma_{ij} w_j - \lambda \bar{r}_i - \mu = 0 \ for \ i = 1,2, \dots, n$$

$$\sum_{i=1}^{n} w_i \bar{r}_i = \bar{r}$$

$$\sum_{i=1}^{n} w_i = 1$$

The set of all efficient or minimum-variance portfolios satisfy the above system of equations.

In the presence of restrictions on short selling, we impose an additional inequality constraint that all the weights need to be greater than or equal to zero. In this case, the resulting Markowitz optimisation problem cannot be reduced to the solution of a set of linear equations. Such an optimisation problem with a quadratic objective and equalities and inequalities as constraints is called a quadratic program, which can be solved using computer programs.

The Two-fund Theorem

Portfolios on the minimum-variance set have an interesting property, which is that only two portfolios (i.e. two sets of weights) are required to determine the entire minimum-variance set. This means that all investors seeking efficient portfolios can create one by investing in a linear combination of just two efficient portfolios, in accordance with their risk preferences. This is the two-fund theorem.

To demonstrate the validity of this theorem, consider two sets of solutions (i.e. two sets of efficient portfolios) to the *n+2* linear system which the minimum-variance set satisfies.

$\boldsymbol{w^1} = (w_1^1, w_2^1, \dots, w_n^1), \lambda^1, \mu^1$ and $\boldsymbol{w^2} = (w_1^2, w_2^2, \dots, w_n^2), \lambda^2, \mu^2$

These have different expected rates of return $\bar{r}^1$ and $\bar{r}^2$. A linear combination of these two efficient portfolios also satisfies the system of *n+2* linear equations and is, therefore, an efficient portfolio. This is because

$$\boldsymbol{\alpha w^1 + (1 - \alpha) w^2}$$

is a portfolio whose weights sum up to 1, which satisfies the weight constraints, and because the expected return of this portfolio

$$\alpha \bar{r}^1 + (1 - \alpha) \bar{r}^2$$

results from a weighted sum of the individual returns. The latter satisfies the return constraint. Finally, the above two expressions make the LHS of the following equation equal to zero, as a result of which they also satisfy that equation.

$$\sum_{j=1}^{n} \sigma_{ij} w_j - \lambda \bar{r}_i - \mu = 0 \ for \ i = 1,2, \dots, n$$

The One-fund Theorem

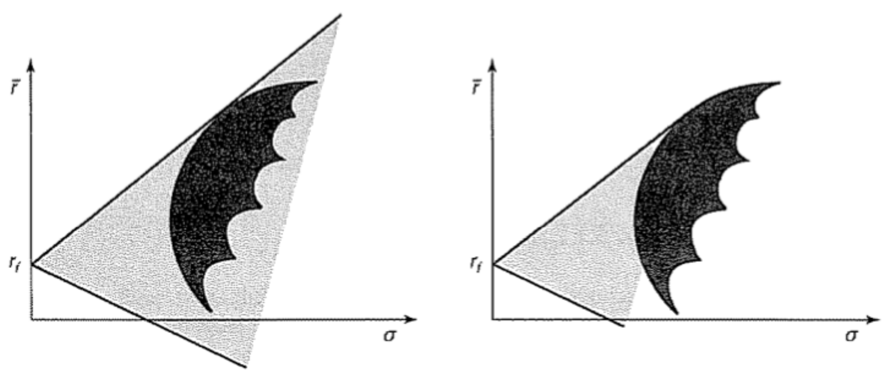
Introducing a risk-free asset reduces the efficient frontier to a straight line, as shown in the diagram below. A risk-free asset has a deterministic rate of return r_f whose volatility $\sigma_f = 0$. So, a portfolio comprising of a risk-free asset and a risky asset in proportions α and $(1-\alpha)$ respectively will have:

$$r_p = \alpha r_f + (1-\alpha)\bar{r}$$

$$\sigma_p = \alpha\sigma_f + (1-\alpha)\sigma$$

The standard deviation simplifies because of absence of correlation between the risky and the risk-free asset. Including the risk-free asset along with *n* risky assets also changes the shape of the feasible set, as shown in the below diagram.

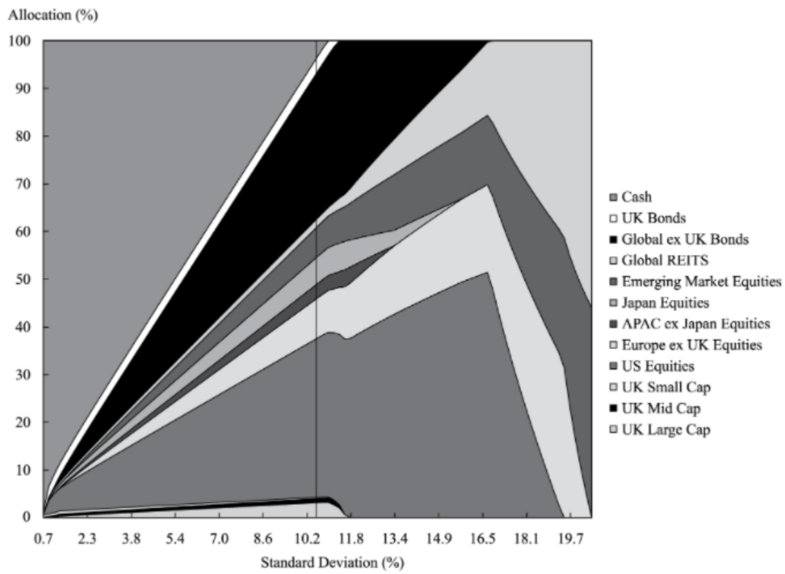
The risk-free asset located on the vertical axis can be combined with every point on the original feasible region with a curve-shaped boundary, to form a portfolio. Since such a portfolio's return and standard deviation are linear combinations of the risk-free and risky assets, it can be represented as a straight line. There is such a line for every point or risky asset on the original feasible region, and their totality results in a triangular feasible region.



The efficient frontier now is simply the highest line of the triangular feasible region. This line is tangent to the original feasible set of risky assets at point F, which is the portfolio of risky assets. Any efficient portfolio can be constructed using this portfolio and the risk-free asset. This is called the one-fund theorem.

Use Cases

The Markowitz model and the efficient frontier are primarily used in practice to develop and set asset allocation policy. This typically begins with an assessment of the investor's risk tolerance, the opportunity set of asset classes he faces, and an estimation of the expected returns, standard deviation and correlation of returns across the asset classes in this set. These inputs are then fed into an optimisation programme which results in optimal asset allocations for various risk levels, as shown below.



Once the investor chooses a level of standard deviation in accordance with their risk tolerance, the Markowitz Mean-Variance optimisation (MVO) provides an elegant solution in the form of the most favourable asset allocation for that risk level.

However, MVO is a single-period analytical framework which cannot incorporate many practical considerations such as the likelihood of an asset allocation meeting certain goals, potential maximum drawdowns, distribution of expected value through time etc. That said, MVO can be complemented with Monte Carlo simulations to address these considerations which are difficult to formulate analytically. Other drawbacks of MVO include:

- Outputs (asset allocations) that are highly sensitive to small changes in the inputs
- Highly concentrated asset allocations in a subset of the available asset classes
- Many investors are concerned about more than the mean and variance of returns, the focus of MVO
- Although allocations may appear diversified across assets, the sources of risk may not be diversified
- Most portfolios exist to pay for a liability or consumption series, and MVO allocations are not directly connected to what influences the value of the liability or the consumption series
- Does not take account of trading/rebalancing costs and taxes

The first two criticisms have been addressed through a reverse optimisation approach, which solves for expected returns, given asset allocation, covariances and the risk aversion coefficients. It is most common to use weights based on market capitalization of the asset classes in the opportunity set for this process. Allocations based on reverse optimisation are well-behaved and well-diversified because assets' expected returns are linked to their systematic risk.

However, the above approach does not leave room for the investor’s forecasts on expected returns to be incorporated. The Black-Litterman model overcomes this limitation by providing a technique for altering the returns derived using reverse optimization and still providing diversified, well-behaved allocations.

Finally, MVO forms the backbone of other non-normal optimisation approaches which take into account other measures of risk such as semi-variance, conditional VaR, skewness and kurtosis. That said, it forms a reasonable starting point to set asset allocation policy.

Application in Project

Daily adjusted closing price data is imported for ETFs covering 11 US sectors. Based on this, we analyse the portfolio statistics and efficient frontier for a 2-ETF portfolio (XLK, XLI) and a 3-ETF portfolio (XLK, XLI and XLY) of our choice.

Fig 1

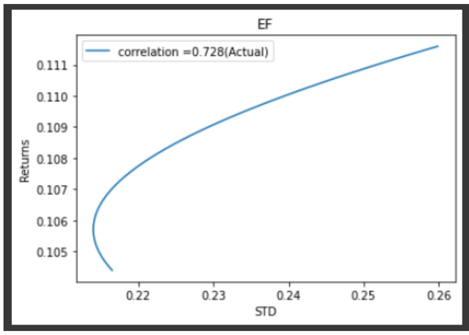


Fig 2

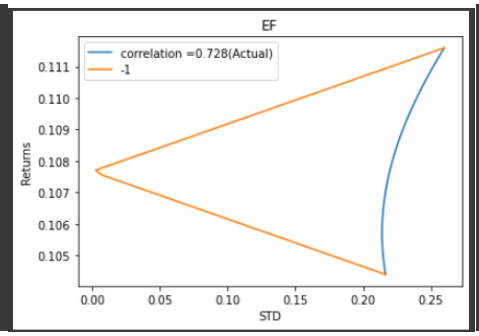


Fig 3

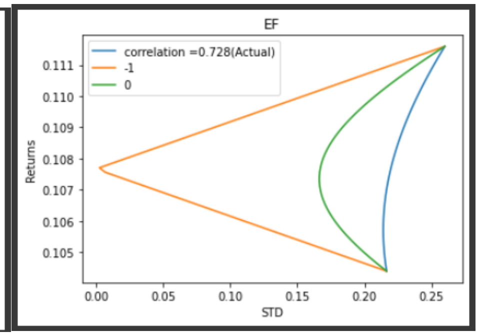


Fig 4

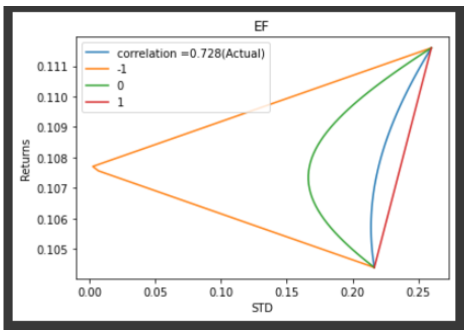


Fig 5

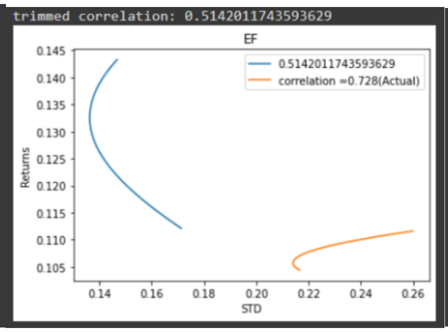
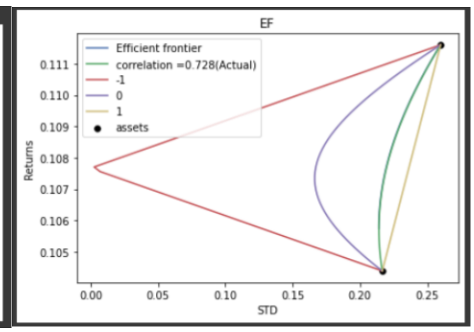


Fig 6



The sectors when compared shows the correlation of 0.72 which is been seen in the 1st figure above. When pretended the correlation between the securities is -1. The shape obtained show the perfect relationship, straight line which can be seen on the 2nd figure. When pretended the correlation between the securities is -1, zero-correlation EF is located between the actual-correlation 0.728 EF and the perfectly negative correlation EF seen on figure 3rd. When pretended the correlation between the securities is 1, negative correlation EF is connected by a perfect correlation of EF which can be seen on the figure 4th. When the data was trimmed by 5% and the correlation and EF was obtained again, the correlation of the actual was seen much better than the correlation of the trimmed which can be seen on figure 5th. The figure 6th shows the efficient frontier along with actual correlation, 0, 1 & -1.

Robust Techniques for Efficient Frontier Estimation – Covariance Shrinkage and the Critical Line Algorithm

Overview

As discussed in the previous sections, MVO has many drawbacks, some of which are addressed in this part. The first technique to enhance MVO discussed here is covariance shrinkage, which makes the estimation of the efficient frontier more robust to small changes in inputs. The second technique covered is the Critical Line Algorithm, which is a numerical approach to derive the entire efficient frontier in a portfolio optimisation problem under linear equality and inequality constraints.

Covariance Shrinkage

Since MVO is highly sensitive to even the smallest deviations in inputs, it is essential that only high-quality inputs are fed into the model. We typically use the historical mean returns and sample covariances as our estimates of expected returns and risk.

Although the sample covariance matrix is an unbiased estimator of covariance, it is an unstructured estimator which often has coefficients with a high amount of estimation error. This is even more true when there are fewer training examples than the number of stocks. Because the MVO model is highly sensitive to these, covariance shrinkage can be employed to move these extreme values towards the centre.

The basic idea behind shrinkage is to switch from an unstructured estimator S, like the sample covariance matrix, to a weighted average of it and some highly structured estimator F, called the shrinkage target.

$$\delta F + (1-\delta)S, 0 \leq \delta \leq 1$$

The weight allocation to the unstructured vs. structured estimators is determined by the shrinkage coefficient δ. Ledoit and Wolf (2004) use the constant correlation model where all pairwise correlations are set equal to the average of all sample correlations. Accordingly, the components of the shrinkage target F is given by:

$$f_{ii} = s_{ii} \qquad f_{ij} = \bar{r}\sqrt{s_{ii}s_{jj}}$$

where s_{ij}^2 is the sample covariance between two assets.

The shrinkage coefficient δ is estimated using the following:

$$\hat{\kappa} = \frac{\hat{\pi} - \hat{\rho}}{\hat{\gamma}}$$

Consider T to be the number of datapoints and y_{it} to be the returns of the ith security at time t. $\hat{\pi}$ is the sum of asymptotic variances of the entries of the sample covariance matrix scaled by $\sqrt{T}$.

$$\hat{\pi} = \sum_{i}^{N}\sum_{j}^{N} \hat{\pi}_{ij} \text{ with } \hat{\pi}_{ij} = \frac{1}{T}\sum_{t=1}^{T} \left\{(y_{it} - \bar{y}_i)(y_{jt} - \bar{y}_j) - s_{ij}\right\}^2$$

$\hat{\rho}$ is the sum of asymptotic covariances of the entries of the shrinkage target with the entries of the sample covariance matrix, scaled by $\sqrt{T}$.

$$\hat{\rho} = \sum_{i=1}^{N} \hat{\pi}_{ij} + \sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} \frac{\bar{r}}{2}\left(\sqrt{\frac{s_{jj}}{s_{ii}}}\hat{\theta}_{ii,ij} + \sqrt{\frac{s_{ii}}{s_{jj}}}\hat{\theta}_{jj,ij}\right)$$

where

$$\hat{\theta}_{ii,ij} = \frac{1}{T}\sum_{t=1}^{T} \{(y_{it} - \bar{y}_i)^2 - s_{ii}\}\{(y_{it} - \bar{y}_i)(y_{jt} - \bar{y}_j) - s_{ij}\}$$

$\hat{\gamma}$ estimates the misspecification of the shrinkage target:

$$\hat{\gamma} = \sum_{i=1}^{N}\sum_{j=1}^{N} \left(f_{ij} - s_{ij}\right)^2$$

The optimal shrinkage constant is given by $\frac{\hat{\kappa}}{T}$, but because this value can be greater than 1 or less than 0, we re-formulate it as follows:

$$\hat{\delta} = \mathrm{m} \left\{ 0, \mathrm{m} \left\{ \frac{\hat{\kappa}}{T}, 1 \right\} \right\}$$

Critical Line Algorithm

As stated previously, optimisation problems with a quadratic objective and equalities and inequalities as constraints do not have an analytic solution, which is why we need numerical approaches to optimise these functions. In the context of portfolio optimisation, the Critical Line Algorithm (CLA) is one such approach.

CLA was developed by Harry Markowitz to optimize general quadratic functions subject to linear inequality constraints. CLA can also be used in a Bayesian portfolio optimisation framework which also leads to a quadratic objective function. However, for some portfolio optimisation problems which do not have quadratic objectives (the Sharpe ratio Efficient Frontier framework, for example), the CLA cannot be used.

Nevertheless, CLA is the only algorithm specifically designed for inequality-constrained portfolio optimization problems, which guarantees that the exact solution is found after a given number of iterations, and derives the whole efficient frontier. This is quite favourable when compared with gradient-based algorithms which may converge to a local optimum depending on how they are initialised and are sensitive to boundary constraints.

Procedure Overview

The algorithm is initialised by finding the first turning point. This consists of the smallest subset of assets which offer the highest expected return such that the sum of their upper boundaries is greater than or equal to unity. We then compute a sequence of turning points, each with successively lower expected return than the previous point. The estimation of each turning point results in a free asset, and the transition from one turning point to the next requires that one element is either added or removed from the subset of free assets. This recursion of adding or removing one asset from F continues until the algorithm determines that the reduction in optimal expected return cannot is minimised and cannot be further reduced.

Associated with each turning point are two parameters λ and γ, both of which can be computed as follows:

$$\lambda = \frac{1}{C}\left[(1 - 1'_{n-k}\omega_B + 1'_k \Sigma_F^{-1} \Sigma_{FB} \omega_B)(\Sigma_F^{-1} 1_k)_i - (1'_k \Sigma_F^{-1} 1_k)(b_i + (\Sigma_F^{-1} \Sigma_{FB} \omega_B)_i) \right]$$

where

$$C = -(1'_k \Sigma_F^{-1} 1_k)(\Sigma_F^{-1} \mu_F)_i + (1'_k \Sigma_F^{-1} 1_k)(\Sigma_F^{-1} 1_k)_i$$

$$b_i = \begin{cases} u_i & \text{if } C_i > 0 \\ l_i & \text{if } C_i < 0 \end{cases}$$

and

$$\gamma = -\lambda \frac{1'_k \Sigma_F^{-1} \mu_F}{1'_k \Sigma_F^{-1} 1_k} + \frac{1 - 1'_{n-k}\omega_B + 1'_k \Sigma_F^{-1} \Sigma_{FB} \omega_B}{1'_k \Sigma_F^{-1} 1_k}$$

where

$$\omega_F = -\Sigma_F^{-1} \Sigma_{FB} \omega_B + \gamma \Sigma_F^{-1} 1_k + \lambda \Sigma_F^{-1} \mu_F$$

Application in Project

Daily adjusted closing price data is imported for ETFs covering 11 US sectors. Based on this, we analyse the portfolio statistics and efficient frontier for a 2-ETF portfolio (XLK, XLI) and a 3-ETF portfolio (XLK, XLI and XLY) of our choice. We then extend this analysis to all 3-ETF portfolios which can be formed from our 11-ETF dataset. Thereafter, the economic indicator (LEI, CEI, LAG) and positional (long + / short -) categories for each ETF in each portfolio are specified, to look for indicator buckets which tend to be positively associated with returns. Finally, we construct a new portfolio from the first 3 principal components of our 11-dimensional dataset and compare its return performance over 2019 and 2020 with our 3-ETF portfolios.

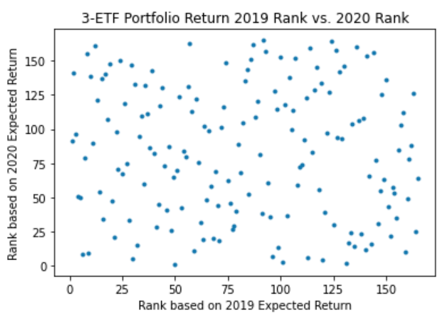
Efficient Frontier Estimation

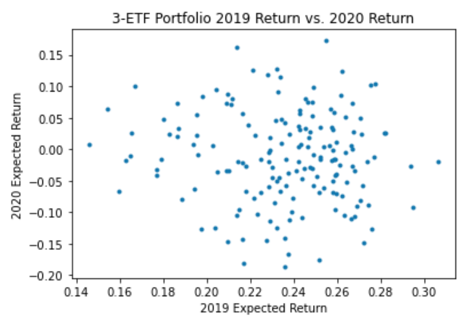
The Critical Line Algorithm (CLA) along with a covariance shrinkage has been used to estimate efficient frontiers for all 165 3-ETF portfolios which can be constructed from our 11-ETF dataset. This has been done for 2019 returns CLA was developed by Harry Markowitz to optimise general quadratic functions subject to linear inequality constraints. CLA solves any portfolio optimisation problem that can be represented in such terms, like the standard Efficient Frontier problem.

Once these are estimated, we assume volatility to be 0.08% and find the expected return and ETF weights for the portfolio lying on each of the 165 efficient frontiers. Since the efficient frontier is given as a set of discrete points instead of an equation, we use the expected return and weights available for the volatility level nearest to our assumed volatility level. The resulting portfolios are our trained portfolios.

Predictive Power of Trained Portfolios

We then apply the weights in the above table to estimate the 2020 expected return for all 165 3-ETF portfolio combinations as a weighted average of the individual ETFs' 2020 expected returns (calculated based on 2020 daily return data).





The above two plots show that the expected return/ rank for 2019 is not a good predictor of the expected return/ rank for 2020 respectively.

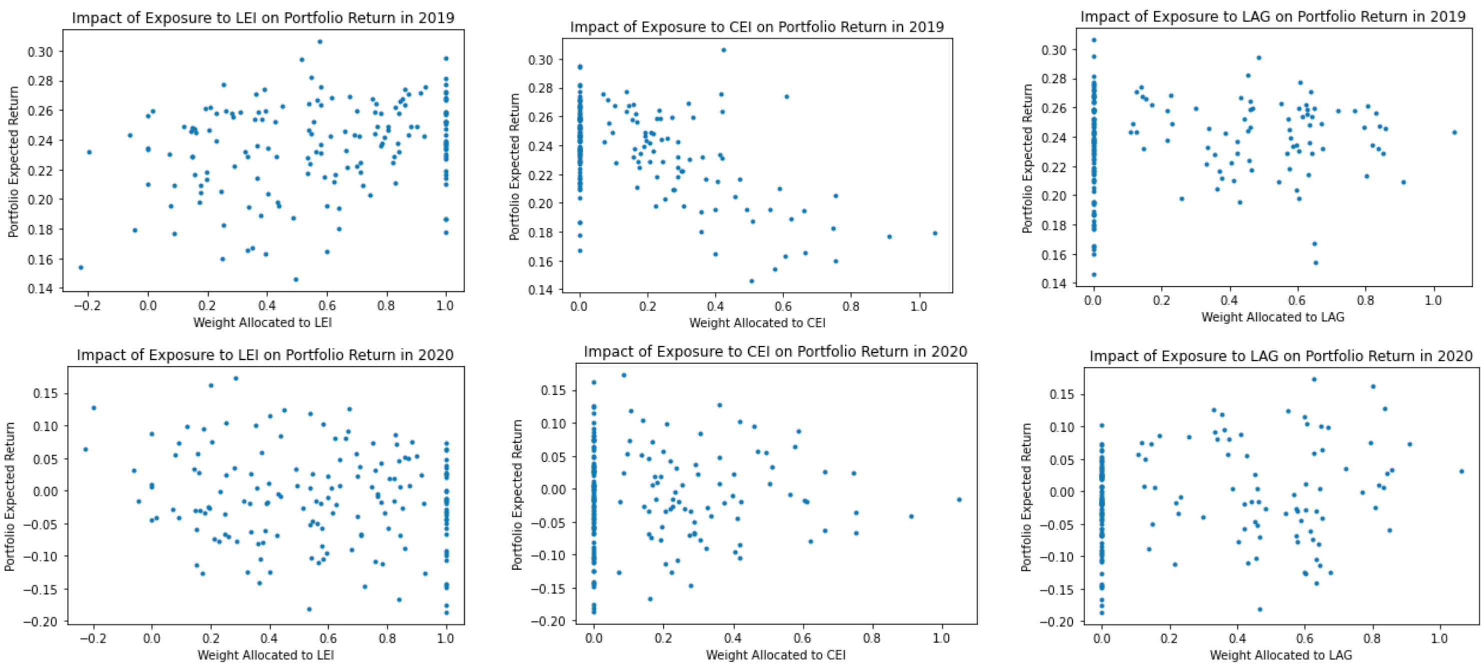
3-sector Portfolios – Analysing Impact of Exposure to LEI, CEI and LAG Indicators on Portfolio Returns

Using the 3 supervised learning models (Linear, Lasso regressions and Regression tree) used in submission 1, we placed each of the 11 ETFs into the 3 economic indicator buckets (LEI, CEI, LAG). With this, we specify the indicator (LEI, CEI, LAG) and positional (long + / short -) categories for each ETF in each of the 165 portfolios.

Thereafter, 27 unique indicator and positional combinations resulting from the above categories are identified, returns for each of these are aggregated and then ranked for 2019 and 2020. However, this does not reveal much about which indicators tend to perform better.

So, we calculate the weights which each portfolio attributes to LEI, CEI and LAG. This is done by looking at the indicator bucket for each asset in the portfolio, and assigning the weight for that asset to its indicator bucket. If more than one asset in the portfolio are assigned to the same indicator bucket, the weights for those two assets are summed and the result is assigned to the indicator bucket. If no asset is assigned to an indicator bucket, the weight of that bucket in the portfolio is taken to be 0.

We then visualise how the 2019 and 2020 expected returns evolve as the weight allocated to a certain indicator bucket increases.



- Increasing a portfolio's exposure to LEI tended to increase returns over 2019 and decrease returns over 2020
- Increasing a portfolio's exposure to CEI tended to decrease returns over 2019. However, the effect of CEI on 2020 returns appears to be ambiguous upon visual inspection
- Increasing a portfolio's exposure to LAG appears to have a marginal negative impact on 2019 returns and a relatively stronger positive impact on 2020 returns

Efficient Frontier Estimation under Constraints

Overview

In the absence of any constraints on expected return or variance of the portfolio or on weights of the various asset classes, optimal portfolios can be determined using an analytical closed-form formula. Although mathematically elegant, the resulting portfolio may often require heavy use of leverage to finance its holdings and may be of little use in the real world to several investors.

This analytic calculation of the optimal portfolio weights is not possible in the presence of constraints, as a result of which one has to resort to numerical methods for derivation of the efficient frontier. However, introducing such constraints results in a frontier which is customised to the special needs of the investor and, therefore, provides much more utility.

The Approach

To illustrate how the efficient frontier changes under different constraints, we use the fictitious data from Elton and Gruber (1995) Chapter 6. This is shown below.

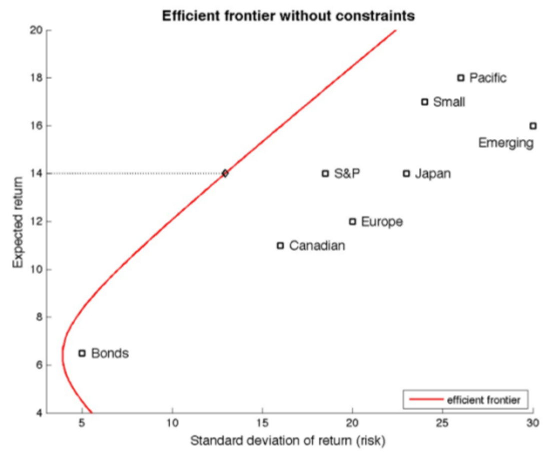
	S & P	Bonds	Canadian	Japan	Emerging Markets	Pacific	Europe	Small Stocks
Expected return	14.0	6.5	11.0	14.0	16.0	18.0	12.0	17.0
Standard deviation	18.5	5.0	16.0	23.0	30.0	26.0	20.0	24.0

S&P	1	.45	.7	.2	.64	.3	.61	.79
Bonds		1	.27	-.01	.41	.01	.13	.28
Canadian			1	.14	.51	.29	.48	.59
Japan				1	.25	.73	.56	.13
Emerging					1	.28	.61	.75
Pacific						1	.54	.16
Europe							1	.44
Small								1

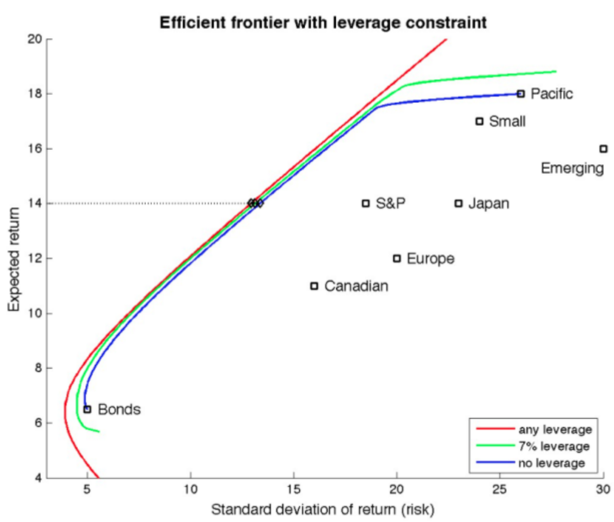
The efficient frontier without constraints which results from the above data is shown alongside. Although unconstrained, this frontier comes with two caveats:

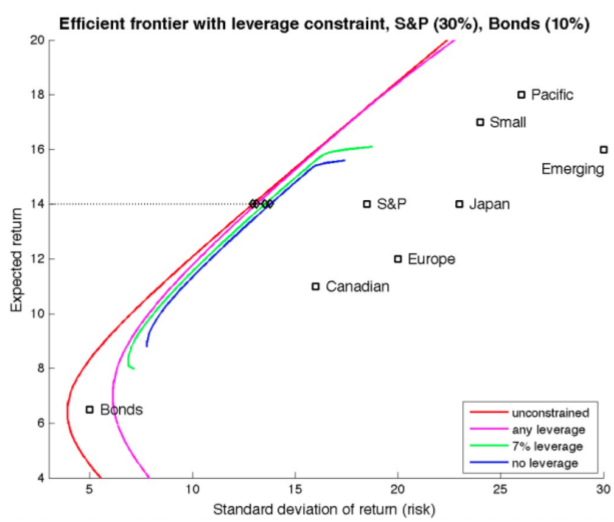
- Optimum portfolios on the efficient frontier are extremely sensitive to the inputs fed into the optimisation. As a result, high quality inputs are paramount for results to be meaningful

- Without constraints, the efficient frontier extends infinitely, which implies that it is theoretically possible to extract any level of return, often by making use of a high degree of leverage



We now separately discuss the impact of introducing two kinds of constraints on the shape of the efficient frontier: a constraint on the level of leverage and a constraint on positions in selected asset classes.





As shown in the diagram on the left, constraining leverage levels shifts the efficient frontier inwards and makes it a finite curve. An interesting point to note is that throughout most of the interesting range of returns (i.e. 8-17%), the decline in return resulting from constraining leverage levels, for a given level of risk, is quite minimal when compared with the large variation in borrowing levels for the three curves.

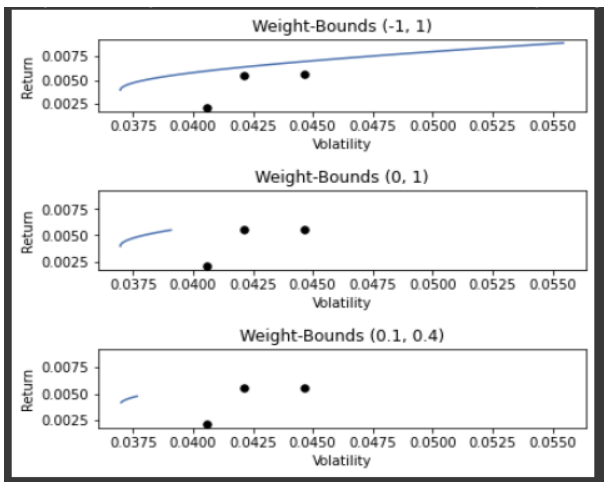
In the diagram on the right, the impact on the efficient frontier from enforcing minimum weight requirements for two asset classes, viz. S&P500 and Bonds is shown. Introducing this constraint narrows down the range of returns available on the efficient frontier. Nevertheless, optimal portfolios on this frontier are not far from those on the original unconstrained frontier for the narrowed return range.

Therefore, it helps to incorporate as many practical constraints as possible on the underlying portfolio model in order to obtain portfolios that are meaningful to the investor.

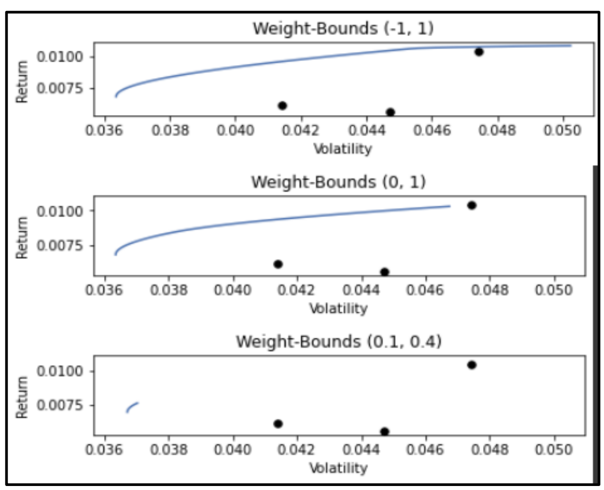
Application in Project

We wrap up our ETF research by examining the 165 portfolio combinations we created earlier, drawing three sample 3-ETF portfolios, and comparing the portfolio's efficient frontier to an efficient frontier under long-only constraints and an efficient frontier under box constraints of 10% minimum weight and 40% maximum weight. The nine efficient frontiers are depicted here, along with the portfolio combination that was used to create the frontier.

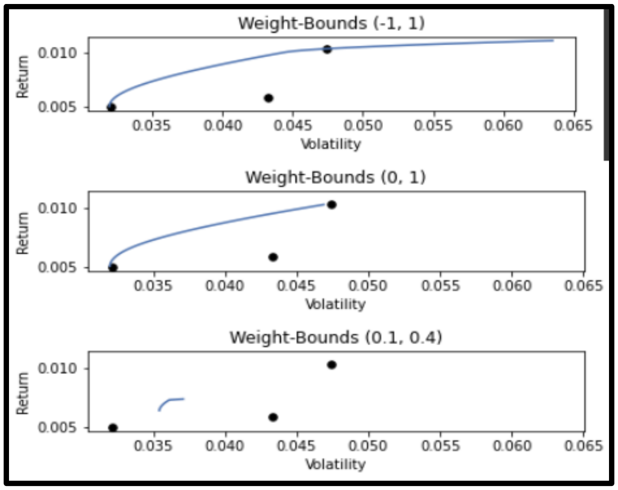
IYR, IYZ, XLB



XLB, XLK, XLU



XLI, XLK, XLP



The EF with no constraints provides the widest range of returns and volatility. Because negative holding permits for short holdings, this is the case. Volatility rises in tandem with greater rewards. The capital required for a no-constraint portfolio should be substantially larger than for a long-only portfolio due to the leverage given by short holdings. The increasing cost of borrowing capital should be considered into the portfolio construction of the asset manager. This indicates that instead of a constant risk-free interest rate, a higher borrowing rate may be necessary in the optimization. The long-only EF is roughly the left part of the no-constraint EF because the EF return cannot exceed the largest return of the underlying three ETFs. For any goal volatility, the best feasible return for long-only EF is somewhat lower than the highest potential return for no-constraint EF (x-axis). The min-max (0.1-0.4) EF has a substantially narrower range of return volatility due to the set range of weights. The use of weight limitations could lead to a more stable EF portfolio. This could be a beneficial feature given the ambiguity around the input assumptions (expected return and covariance).Weight bounds may be required due to fiduciary responsibility, which requires the asset manager to advise the client on investment strategy while taking into consideration any investment constraints the client may have.

Conclusion

To conclude, we found in the first part of the project that most models placed each sector ETF to similar economic indicator buckets, although some discrepancies were noted.

Thereafter, in the 2-sector portfolio analysis covered in the second part of this project, the actual correlation was much better than trimmed data correlation. The correlation relationship had been validated comparing with 0, 1 and -1 correlation. In the 3-sector portfolio analysis, we found that expected return/ rank for 2019 was not a good predictor of the expected return/ rank for 2020 respectively. LEI, CEI and LAG tended to have different effects on returns over 2019 and 2020, so no one indicator clearly stood out. Finally, the return of the PCA portfolio for both 2019 and 2020 was below the average return for the 3-sector portfolio.

Lastly, the third part revealed that the efficient frontier under no constraints provided the widest range of returns achievable. Nevertheless, imposing constraints in accordance with the investor’s specific situation produced more relevant results and portfolios which did not compromise much too much on returns relative to their unconstrained counterparts.

References

Ganti, A. (2022). “What Is the Efficient Frontier?” Investopedia www.investopedia.com/terms/e/efficientfrontier.asp

Jaadi, Z. (2021). “A Step-by-Step Explanation of Principal Component Analysis (PCA).” Built In builtin.com/data-science/step-step-explanation-principal-component-analysis

Martin, RA. (2018). User Guide - PyPortfolioOpt 1.5.1 Documentation pyportfolioopt.readthedocs.io/en/latest/UserGuide.html

Xu, W. “What's the Difference between Linear Regression, Lasso, Ridge, and ElasticNet?”

Luenberger, D. (1998). Investment Science. Oxford University Press

Ledoit, O., & Wolf, M. (2004). Honey, I Shrunk the Sample Covariance Matrix. The Journal of Portfolio Management, 30(4), 110–119

Bailey, DH., Prado ML. (2013). An Open-Source Implementation of the Critical-Line Algorithm for Portfolio Optimization. Algorithms