Using Clustering Algorithms for Investment Due Diligence and Portfolio Construction

Navigating the Tricky Waters of Building an Investment Portfolio

Building a solid investment portfolio is no easy task, especially when we dive into the world of portfolio optimization - trying to make decisions that give the best returns while keeping risks low. Traditional methods often lean on historical data to estimate future returns and how different investments behave relative to one another (correlations). But here's the catch: financial markets are not always predictable and relying heavily on past performance can sometimes steer us wrong. Small changes in expected returns or shifts in how assets move in relation to one another can shake our portfolio's stability, exposing it to unexpected risks. We can attempt to improve optimization with sophisticated models, intricately designed to forecast future returns and correlations. However, they seldom align with actual future occurrences. They encapsulate myriad variables and leverage advanced algorithms to mitigate the problems associated with traditional portfolio optimization. However, the eventualities of global economies, policy changes, and unexpected market disruptions invariably usher unanticipated variances, unearthing portfolio vulnerabilities.

This is where hierarchical clustering can step in and be a useful tool. It doesn't just look at past performance but can also group investments together based on a wider range of criteria. This can help build a portfolio that's not only grounded in understanding past behaviors but also resilient against the unpredictable ebbs and flows of the market.

Understanding Investment Behaviour

At the heart of successful portfolio construction lies a deep understanding of how asset classes and individual investments evolve and behave over time. While the financial world is replete with metrics and numbers that capture snapshots of investments at specific moments, these point-in-time estimates—whether they be of style, correlations, or returns—are but momentary aspects in the grand tapestry of an investment's journey. These static measures, while informative, are akin to studying a single frame of a movie, missing out on the unfolding narrative and the dynamic interplay of its characters. Portfolio managers and registered investment advisors have a duty to fully understand the products that they use for clients. Do point in time metrics really demonstrate understanding of an investment's risks?

The greatest risk that a portfolio manager faces are not tail risks or unknown future events; these will occur and normally cannot be avoided. Risk management techniques will help mitigate the effects of these events, but all investors know these events will occur. Rather, the greatest risk portfolio managers face are the ones they believed they understood and then the opposite occurred. Those are risks that can lose clients, or worse, result in a lawsuit.

So then how can we grasp a better understanding of investment behaviour?

Understanding factor exposures thus becomes incredibly important as this will provide a much better mechanism to understand behaviour. But it is insufficient to look at point in time estimates of factor exposures. Understanding the evolution of those factors is what will guide you in understanding the expected behaviour of an investment and ultimately build better portfolios.

Using Machine Learning to Improve our Factor Models

Once we have constructed our factor models, we can use various machine learning techniques to improve our models, thus producing more reliable estimates of our factor exposures over time. Feature selection techniques will help eliminate unneeded factors from the models and at the same time, reduce correlations between factors. Once these factors are removed, we can rerun all the factor models in successive time periods, and we should see improvements in the results.

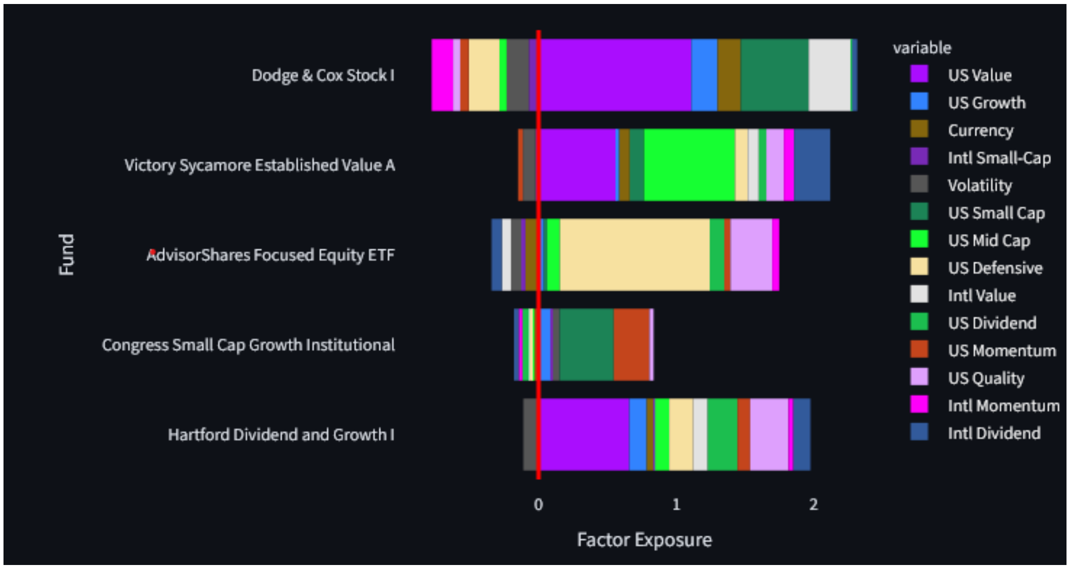
Once these models are completed, we can take investments, examine their exposures, and then answer the following questions:

- Does the investment provide the factor/risk exposures we are looking for?
- Does the exposure match the manager's stated style and objective?
- How similar are the exposures of funds that have the same stated objective?

This process will help uncover the needed details to perform investment due diligence. And given the quagmire presented by current Know Your Product requirements, getting a firm understanding of investment risk exposure over time is of extreme importance. This method will help you not only assess the style of an investment, but because we calculated it over multiple time periods, it considers how these exposures have changed over time.

For example, assume we were analyzing the US Equity universe, and we have 150 funds and ETFs that we are looking at. We create factor models for each fund, and we can then visually compare the exposures of funds of interest to ensure that they fit our criteria. This could also be used to assess how a new fund that is being considered for inclusion into a model compares with existing holdings or to identify new candidates for a model portfolio.

The following figure shows the factor exposures for 5 US equity mutual funds.



The factor model shows the exposures to the key factors that explain the variation of the 150 funds used in this analysis. For example, the AdvisorShares Focused Equity ETF has US Defensive as its main factor exposure. So, we would expect this fund to be more conservative and lower volatility and could be a good candidate for our US equity defensive line in our model portfolio. We could then combine this fund with a more aggressive one, or a small-cap fund for our high-risk budget line in our portfolio.

Using this method would absolutely demonstrate Know your Product. You can confirm the manager's stated style, match it to what you are looking for, and then be able to thoroughly explain why this manager was included in your model.

Constructing Cohesive Portfolios with the Aid of Hierarchical Clustering

If we had a universe of 150 funds, visually comparing them is not ideal. This is where hierarchical clustering plays a central role in our investment due diligence and portfolio construction pathway.

So far, we have performed investment due diligence using our improved factor models. Our goal is to build a diversified portfolio. Normally, we would not use only one investment per asset class. We might want two to three managers for US equities, potentially sub-dividing further by market capitalization.

As an analogy, we can think of this as two managers playing in the same sand box but using different toys to build a similar sandcastle. In our portfolio, we might have two managers benchmarked to the S&P 500 and attempt to beat it with different styles and strategies. For example, one can be a dividend growth manager who believes that companies that consistently increase dividends over time will deliver strong alpha in the long run. Another manager believes that allocating tactically to different sectors, growth, and value, and even adding in some non-US equities will deliver the needed alpha. Combining these two in a portfolio would make sense; you get style diversification that has been verified by our factor models in step one.

Thus, our task is to find the optimal combination of investments that increases diversification.

The next step is to use the factor exposures for each fund as inputs to a hierarchical clustering algorithm. What this means is that we can assess how similar or dissimilar funds are from one another based on their factor exposures. Now, say we have 16 factors that can explain all the variation within the returns of our fund universe. The human brain is not capable of mapping the distances between investments over 16 dimensions. But machine learning can, and this is exactly what hierarchical clustering can do for us.

Our hierarchical clustering algorithm will group investments together based on their exposures, and by amalgamating all factors, it will assess which funds are managed in a similar (or dissimilar) way. This means that we can have one summary number that encapsulates fund similarity based on their management styles and factor exposures. With a universe of 150 funds, this equals 22,500 distance scores!

Returning to our example, let's say we currently have AdvisorShares Focused Equity in our model, and we want to complement it with another US equity fund, but preferably with different factor exposures so that we can improve upon diversification and not be overly exposed to the same factor or risk exposures.

The following table shows which funds are most similar or dissimilar to AdvisorShares Focused Equity based on the factor exposures.

Most Similar Funds

	AdvisorShares Focused Equity ETF
Dearborn Partners Rising Dividend I	0.3518
VictoryShares Dividend Accelerator ETF	0.5739
T. Rowe Price Dividend Growth I	0.5782
Vanguard Dividend Appreciation ETF	0.7731
American Beacon The London Co Inc Eq R5	0.8753

Most Dissimilar Funds

	AdvisorShares Focused Equity ETF
Schwab S&P 500 Index Fund	3.1053
Vanguard S&P 500 ETF	3.1867
Vanguard Total Stock Market Index Fund Admiral Shares	3.2044
SPDR S&P 500 ETF	3.2048
Vanguard Total Stock Market ETF	3.4610

The clustering algorithm recommends combining our fund with an S&P 500 index fund. We can also see that the funds offering the least amount of diversification would be other dividend funds. This makes perfect sense given the factor exposures seen in Figure 1.

This can help us choose which fund to complement our existing holding because we know it has different factor exposures. We can repeat this process for each of our target asset classes until we have built a portfolio where we have multiple managers with the same objective but are going about achieving it with different tools.

Summary

Using correlations to build portfolios presents many problems. Two US Equity managers will undoubtedly be highly correlated simply because they are benchmarked to the S&P 500. But this does not mean that they have similar strategies or risk exposures, and they might still offer diversification benefits. Correlations will not be able to help identify this. Factor models, refined by machine learning, and performed over multiple time periods provides greater clarity into management styles over time. This would provide us with a powerful due diligence mechanism that would also go above and beyond all Know Your Product requirements.

We can use these factor exposures as inputs to a hierarchical clustering algorithm and construct a diversified portfolio. All without having to rely on expected returns and correlations that are fraught with estimation errors and pitfalls.