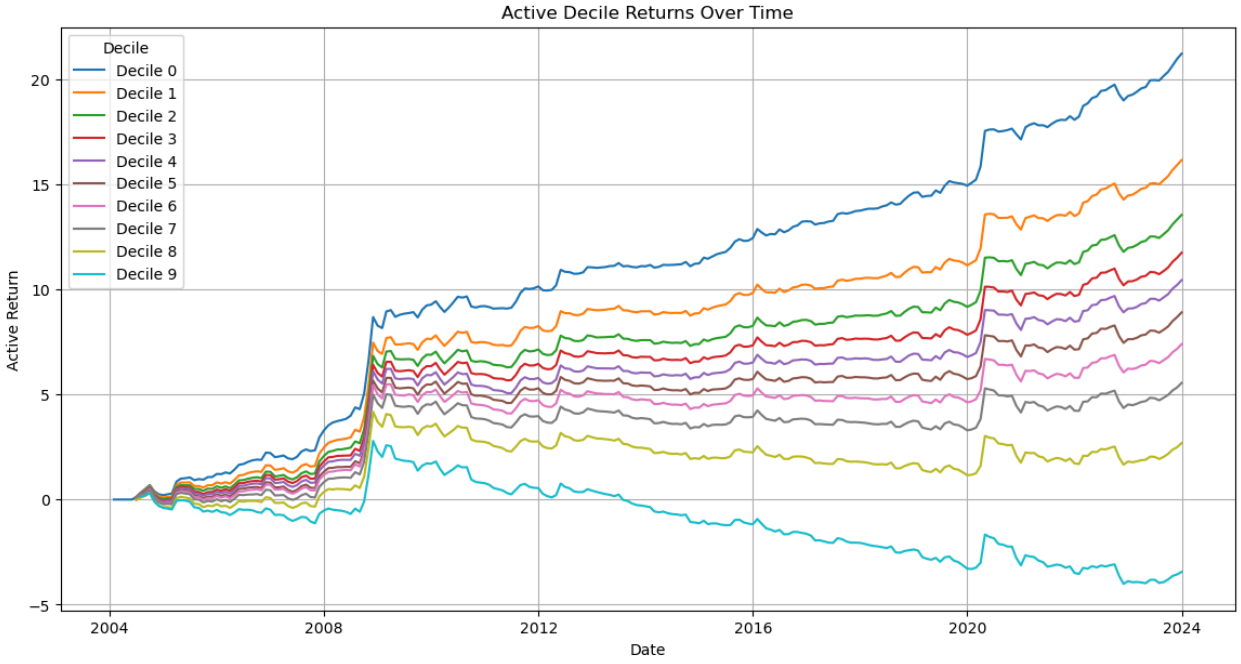
Findings from research on Weighted PCA Algorithm to identify orthogonal factors

- By Pavan Mirla

Key Findings



Active Decile Returns Over Time: This graph displays the cumulative returns for ten portfolios, labeled Decile 0 through Decile 9, created using a weighted PCA approach. Each portfolio is constructed by weighting principal component scores with their corresponding risk premiums to forecast expected returns over time. Decile 0 represents the portfolio with the highest cumulative returns as per this methodology

There are similarities between the Weighted PCA approach and the Arbitrage Pricing Theory (APT), as both are multifactor models used to determine asset prices based on several macroeconomic factors or theoretical market indices.

Approach

Standardize the Data:

To ensure that each variable contributes equally to the analysis.
Process: Subtract the mean and divide by the standard deviation for each variable.

Principal Component Analysis (PCA):

To reduce the dimensionality of the dataset, while preserving as much variance as possible.
Process: Identify the principal components (PCs) which are orthogonal (uncorrelated) linear combinations of the original variables.

Determine the Number of Components:

To choose how many PCs to use in further analysis.
Process: Often decided based on the explained variance, using a scree plot or a predefined threshold (like 85% of variance).

Calculate PC Scores:

To transform the original variables into these new, uncorrelated factors (PCs).
Process: Multiply the original standardized data by the PCA loading vectors (eigenvectors).

Regression Analysis:

To establish the relationship between the PCs and the asset returns.
Process: Regress the asset returns on the PC scores to get coefficients that show the sensitivity of the returns to the PCs.

Weighting the PCs:

To understand the combined effect of all PCs on asset returns.
Process: Multiply each PC score by its corresponding regression coefficient to get weighted scores.

Aggregation of Scores:

To combine the information of all PCs into a single score that can rank assets.
Process: Sum the weighted scores across all PCs for each asset.

Ranking and Portfolio Formation:

To select assets based on their aggregated scores for investment.
Process: Rank assets based on their aggregated scores and form portfolios by grouping them into deciles or quantiles.

Performance Evaluation:

To assess how well the portfolios perform over time.
Process: Calculate the returns of each portfolio and evaluate their performance through metrics like cumulative returns, Sharpe ratio, etc.

How to quantify our expectations of returns based on the portfolio's exposure to different risk factors and their corresponding premiums.

$$R_i - R_f = \alpha_i + \beta_{i1}F_1 + \beta_{i2}F_2 + \ldots + \beta_{in}F_n + \epsilon_i$$

A traditional factor model showing excess returns driven by various factors plus noise.

$$R_i - R_f = \alpha_i + \beta_{i1}PC_1 + \beta_{i2}PC_2 + \ldots + \beta_{in}PC_n + \epsilon_i$$

A modernized model using principal components to capture excess returns.

$$\mathrm{E}(R_i) = R_f + \beta_{i1} \cdot \mathrm{RP}_1 + \beta_{i2} \cdot \mathrm{RP}_2 + \ldots + \beta_{in} \cdot \mathrm{RP}_n$$

Expected return for a stock; risk-free rate plus weighted risk premiums.

$$\mathrm{E}(R_p) = \sum (w_i \cdot \mathrm{E}(R_i))$$

The portfolio's expected return is the sum of weighted individual expected returns.

R_i represents the return of the i^{th} asset.
R_f is the risk-free rate, typically the return of a government bond.
α_i is the asset's unique return not explained by the model.
β_{ij} shows how sensitive the asset's return is to the j^{th} factor or principal component.
F_j or PC_j denotes the j^{th} factor or principal component.
RP_j is the risk premium associated with the j^{th} factor or principal component.
ϵ_i is the error term, accounting for model inaccuracy.
$E(R_i)$ is the expected return of the i^{th} asset.
w_i represents the weight of the i^{th} asset in the portfolio.
$E(R_p)$ is the expected return of the portfolio.

PCA vs. APT Factors:

In APT, returns are explained by a few common risk factors which are believed to affect all securities. PCA identifies these factors empirically by finding the combinations of observed variables that explain the most variance in returns.

The principal components can be considered as proxies for the unobservable factors in APT. Each PC is a factor that captures a specific element of systemic risk.

Regression to Find Risk Premium:

The regression of asset returns on the PCA factors is akin to the factor sensitivity analysis in APT. This step estimates the risk premium associated with each factor.

The coefficients obtained from regression represent the price of risk for each factor. This tells us how much additional return is expected for taking on an additional unit of risk associated with each factor.

Ranking and Portfolio Construction:

Ranking stocks by their weighted score and constructing portfolios is an application of the APT model, which suggests that in the absence of arbitrage, assets will be priced such that their expected returns will compensate for their risk as measured by their loadings on the factors.

Technical Meaning: By constructing portfolios based on weighted PCA scores, investors are implicitly betting on the persistence of the risk premiums associated with the identified factors.