# Context-Aware Alpha Extraction (CAAE) Strategy

**Gabriele Monti** montig@usc.edu

**Juan Herrera** jjh80024@usc.edu

**Fernando Garcia de Llano** fgarciad@usc.edu

## Abstract

This paper presents the Constant Alpha, Alpha Everywhere (CAAE) strategy, a dynamic factor-based investment approach that aims to outperform traditional benchmarks. By implementing a regime-based allocation framework, CAAE adapts to changing market conditions, enabling it to capitalize on opportunities and mitigate risks. The strategy involves constructing 12 factor-based portfolios and dynamically adjusting their weights based on macroeconomic indicators like inflation, interest rates, and market volatility. Through backtesting techniques and analysis, we evaluate the performance of the CAAE strategy and its potential to deliver superior risk-adjusted returns.

# Introduction

The Context Aware Alpha Extraction (CAAE) strategy is a high-quality investment plan that works towards optimizing returns by dynamically adjusting to the ever-changing market conditions. Compared to conventional factor investing, which uses static component weights, our strategy uses a data-driven process to modify portfolio allocations in response to specific current economic KPIs.

The strategy’s goal is to build a diversified portfolio of ten factor-based portfolios that focuses on specific characteristics such as value, momentum, or quality by integrating financial factors with alternative data sources.

A key innovation of CAAE is its ability to allocate weights to these portfolios based on prevailing market regimes. The strategy identifies optimal investment opportunities by analyzing macroeconomic indicators like inflation, interest rates, and market volatility and adjusts its exposure accordingly. This adaptive approach enables the strategy to capitalize on changing market dynamics and mitigate risks.

# Methodology

## Data Preparation

To start the analysis, we used two primary datasets: CRSP and Compustat. CRSP provides us with important stock information, including prices, returns, trading volume, and shares outstanding monthly. On the other hand, Compustat delivers a wide variety of financial metrics, including assets, liabilities, and earnings, that come from quarterly reports. Additionally, we used specific datasets to pull information on Fama French, inflation, interest rates, and VIX.

## Data Cleaning

After preparing the data, we used a meticulous data-cleaning procedure to guarantee our project's results and process accuracy. At first, our Compustat data was a quarterly data frame. To convert it to monthly data, we used forward-filling for each first month in each quarter until the next available quarter date. In order to avoid the look-ahead bias, financial KPIs are lagged by three months. The two datasets were merged using permno (CRSP) and gvkey (Compustat). Lastly, additional datasets that include inflation rates, interest rates, and Fama-French factors are also included in the analysis to provide better context.

## Dataset Merging

The CRSP and Compustat datasets are merged based on permno and gvkey. This integration enables us to combine the rich financial information from Compustat with the basic stock-level data from CRSP. Furthermore, we incorporate supplementary datasets, such as inflation, interest rate, and Fama-French factor data, to enrich our analysis and capture broader economic factors.

## Final Dataset Explanation

The resulting merged dataset constitutes the foundation of our analysis. It incorporates monthly stock-level data, including prices, returns, trading volume, and shares outstanding, coupled with comprehensive financial metrics derived from quarterly reports. This dataset, along with the macroeconomic factors, provides a robust framework for investigating the relationships between firm characteristics, market conditions, and stock returns.

# Portfolio Definition and Construction

## Factor-based portfolio

The investment strategy centers on a factor-based approach. We construct 10 factor-based portfolios specified below. In this approach, we build different portfolios based on specific factors, and we rank all the stocks in our dataset depending on specific metrics of the respective underlying portfolio each month.
Each calculated rank will serve for the month after, meaning that when allocating weights to the final portfolio based on the factors of top performers, we have to consider the ranks of the month before.

For example, 120 total stocks were selected (10 from each portfolio) as the highest ranked in their respective portfolio for the month of December 2021; therefore, if we want to allocate weights for January 2022, we should take into account the values for December 2021 and so on.

Below are the 10-factor portfolios that we constructed.

### 1. Value Portfolio

For the Value portfolio, we used two metrics to rank all the stocks: Book-to-Market ratio (B/M), and Enterprise Multiple (EV/EBITDA).
Book-to-market is calculated as book value of equity over market value. For EBITDA, we approximate it as Operating Income Before Depreciation + Capital Expenditures.

### 2. Momentum Portfolio

The momentum portfolio used 6-12 month returns and 12-24 month returns as metrics to rank stocks from best to worst. The 6-12 month returns were calculated by grouping by PERMNO to ensure that the calculation was done for each stock separately. Finally, we shifted the return column by 6 months to exclude the most recent 6 months of data. The same process was used to calculate the 12-24 month returns, but it was shifted to 12 months instead.

### 3. Size Portfolio

The size portfolio is constructed by raking the Market Value metric, which is built as Market Value = |Price| x Shares Outstanding

### 4. Quality Portfolio

The Quality portfolio is constructed by raking the Quality Minus Junk (QMJ) and Earnings-to-Price Ratio metrics.
Quality Minus Junk is calculated as Shareholder’s Equity over Market Value. We calculated Earnings-to-Price Ratio as earnings over price.

## 5. Profitability Portfolio

For this portfolio, the metrics of interest were the following: Gross profitability, ROA, ROE, and profit margin. Gross profitability was calculated by dividing total sales/total assets. ROA was calculated by dividing net income/total assets. Additionally, ROE was done by dividing net income/shareholder equity. Finally, the profit margin was calculated by dividing net income/sales.

## 6. Earnings Quality Portfolio

Earning quality portfolio is constructed by ranking the stocks by accruals, net operating assets, and cash-flow-to-price ratio. Each was calculated in the following ways:
Accruals = (Net Income (ibq) - Cash Flow from Operations) / Total Assets (atq).
Net operating assets = earnings_quality_df['atq'] - earnings_quality_df['ceqq'].
Cash-flow-to-price-ratio = Cash Flow from Operations (cfo) / Market Value (mv).

## 7. Investment & Growth Portfolio

The Investment & Growth portfolio is constructed by ranking the Asset Growth, CAPX, Growth Rate metrics. Asset Growth is calculated as the percent change of total assets month over month. CAPX is provided to us in the dataset “capxy.” The growth rate is the percent change in sales month over month.

## 8. Liquidity Portfolio

The Liquidity portfolio is constructed by raking the Amihud's Liquidity and High-Volume Premium metrics. Amihud's Illiquidity is defined as |Return| / (Price * Volume). The High-Volume Premium is calculated as the product of Volume and Price.

## 9. Inflation Sensitivity Portfolio

The Inflation Sensitivity portfolio is constructed by calculating the correlation between the inflation rate and returns for each PERMNO, in which PERMNO returns are mostly correlated with inflation.

## 10. Past Returns Portfolio

This portfolio is built based on 1-month, 2-5-month lagged returns, and 6-12-month, 12-24 month returns. The 1-month represents the most recent month's performance. The 2-5-month return is calculated by summing the returns from months 2 to 5 before the current month, achieved by shifting the return column by 1 month and applying a 4-month rolling sum. Similarly, the 6-12-month return sums the returns from months 6 to 12, using a 1-month shift and a 6-month rolling sum. Lastly, the 12-24-month return captures the performance from months 12 to 24, calculated by shifting the returns by 12 months and applying a 12-month rolling sum

## Rankings

Based on the 10 different portfolios summarized in the previous section, we ranked each stock, each month from late 2000 to the end of 2023, based on which stocks performed the best based on the portfolio's respective metrics. For example, in the Value Portfolio, where we used Book-to-Market as a metric, we aggregated each stock on a monthly basis and ranked them in descending order based on which stock achieved the highest Book-to-Market on that given month. This ranking allows us to select monthly stocks based on the respective metric.

## Stock Selection

Similarly, after the stocks are ranked, the top 1% of the stocks are selected. For example, if we ranked 1000 stocks based on the Book-to-Market metric, we would keep the 10 stocks that achieved the highest B/M and choose them based on the calculated ranking.

# Market Regime Assessment

## Macroeconomic Metrics

To assess market conditions and guide strategic decisions, we closely monitor our key macroeconomic metrics: Inflation rate, interest rate, and VIX. The inflation rate, extracted from Consumer Price Index (CPI) data, provides insights into economic growth. In contrast, the interest rate is a solid reflection of the monetary policy and its impact on borrowing costs and investment activity overall. Lastly, the VIX offers real-time investor uncertainty as it is a great resource to measure market volatility.

## Regime Classification

We classify market regimes into distinct categories by analyzing the trends and interactions of these macroeconomic metrics. Here, we explain the conditions and guidelines we followed to classify the market as bullish, bearish, neutral, or volatile using the macroeconomic metrics.

We classify the market as bullish [1,2,3,4] when:

1. VIX $< 20$
2. Monthly Inflation Rate $< 0.002$
3. Monthly Interest Rate $< 0.04$

We classify the market as bearish [5,6,7,8] when:

1. VIX $> 30$
2. Monthly Inflation Rate $> 0.005$
3. Monthly Interest Rate $> 0.06$

We classify the market as neutral [9,10,11] when:

1. VIX between 20 and 30
2. Monthly Inflation Rate between 0.002 and 0.005
3. Monthly Interest Rate between 0.04 and 0.06

We classify the market as volatile [12] when:

1. The VIX spikes, VIX $> 25$, regardless of the Interest Rate and Inflation Rate

## Weight Allocation and Market Classification Logic

Our strategy adapts to the changing market conditions by adjusting the weights allocated to each factor portfolio. Here we explain the specific rules that we follow in order to allocate weights to each of the portfolios depending on what market we're on.

Bullish Market Weights:

- Risk-on portfolios dominate, with higher allocations to:
  - Value (20%): Capitalize on undervalued stocks.
  - Momentum (20%): Leverage the persistence of winners.
  - Size (15%): Smaller firms benefit more in growth periods.
  - Past Returns (10%): Medium- to short-term momentum plays thrive.
- Lower weights to defensive factors like Quality, and Earnings Quality.

Bearish Market Weights:

- Defensive portfolios dominate, with higher allocations to:
  - Quality (20%): Focus on financially stable firms.
  - Profitability (20%): High profitability signals resilience.
  - Earnings Quality (15%): Transparent earnings protect during downturns.
  - Distress Risk (10%): Mitigates exposure to financial collapse.
- Reduced exposure to risk-on factors like Momentum and Past Returns.

Volatile Market Weights:

- Emphasize macro sensitivity and liquidity, with higher allocations to:
  - Macro Sensitivity (20%): Hedge against macroeconomic risks.
  - Liquidity (20%): Ensure flexibility in uncertain times.
  - Distress Risk (15%): Maintain stability.
- Moderate exposure to defensive factors like Earnings Quality and Quality.

Neutral Market Weights:

- Allocate equally across all portfolios to balance risks and capture a wide range of return drivers.

# Backtesting Framework

To evaluate the historical performance, we constructed a backtesting procedure to assess the strategy's past success. This entails examining past data to evaluate essential indicators, including maximum drawdown, Sharpe ratio, and annualized returns. The Sharpe ratio measures risk-adjusted performance, considering return volatility, whereas annualized returns calculate the portfolio's compound annual growth rate. The greatest peak-to-trough drop in the portfolio's value is measured by maximum drawdown. Backtest was run through different time periods.

The backtest simulates the portfolio's performance by applying the defined investment strategy over a historical timeline, using monthly returns data for individual stocks. It operates as follows:

## Portfolio Construction

For each month, stocks are selected and allocated weights based on their inclusion in specific factor portfolios and the prevailing market regime, as determined by the macroeconomic data.

## Data Integration

The selected stock allocations are merged with their respective monthly returns from the CRSP dataset. Any missing return data is handled by assigning a zero return, ensuring no gaps in the calculation process.

## Weighted Returns Calculation

Each stock's return is weighted according to its allocation in the portfolio. The portfolio's total return for a given month is the sum of these weighted returns.

## Performance Metrics

Monthly portfolio returns are aggregated to compute performance metrics, including average monthly returns, standard deviation, cumulative returns, and Sharpe ratio. These metrics provide insights into the portfolio's risk-adjusted performance and overall return profile.

This process ensures that the backtest provides a systematic and repeatable method to evaluate the portfolio's alignment with the underlying investment logic.

# Results and Analysis

## Portfolio Performance

The portfolio demonstrates notable performance variability, excelling in certain months and specific date ranges while underperforming when evaluated over the entire 2000–2023 period. This discrepancy in performance can be attributed to several factors inherent in both the investment logic and external market conditions.

### Exceptional Monthly and Short-Term Performance

In specific months, the portfolio achieves high returns, often aligning with favorable market regimes such as bullish or volatile markets. During these periods, the dynamic weight allocation strategy, which emphasizes risk-on portfolios like value, momentum, and size in bullish markets or macro-sensitive portfolios like liquidity and inflation sensitivity in volatile markets, effectively capitalizes on prevailing market trends. These targeted allocations enable the portfolio to outperform benchmarks during brief market conditions that align with its strategic biases.

### Long-Term Underperformance Over 2000–2023

Despite these successes, the portfolio underperforms when aggregated over the full date range. This may stem from several potential reasons:

Overexposure to Certain Factors: The portfolio may have overemphasized factors that underperformed during extended periods. For instance, value stocks, which historically outperform in some cycles, struggled for much of the post-2008 recovery period, adversely affecting overall returns.

Selection Bias in Portfolios: By selecting only the top-performing stocks in each portfolio, the strategy may inadvertently concentrate risk. This lack of diversification can amplify returns in favorable periods but also exacerbate losses in adverse market conditions.

Macroeconomic Headwinds: External events such as the Dot-com crash, the 2008 financial crisis, and the COVID-19 pandemic caused significant market upheavals. These shocks may have disproportionately impacted the portfolio's allocations, particularly if the strategy was overexposed to cyclical or high-volatility assets during downturns.

Rebalancing and Turnover Costs: Frequent reallocation based on monthly performance incurs transaction costs, which may erode gains over the long term. This factor is particularly impactful over a two-decade span with consistent monthly rebalancing.

# Future Improvements and Recommendations

The current analysis, while comprehensive, identifies several limitations and opportunities for future enhancement. These are primarily centered around data quality, methodological refinements, and mitigation of biases inherent in the present approach.

## Addressing Look-Ahead Bias

Market regime classification in the current framework is based on contemporaneous monthly data. This introduces a look-ahead bias, as the regime classifications are assumed to be known without accounting for real-world delays in data availability. In future iterations, we will use lagged macroeconomic indicators to align classifications with actual market conditions at the time of investment decisions. This correction will enhance the robustness and realism of the backtesting process.

## Data Completeness and Merging Challenges

The WRDS CRSP monthly data reveals notable gaps, with multiple permnos missing data for specific dates. These discrepancies will be addressed by ensuring that the merging of CRSP and Compustat datasets occurs only on dates with available data for both sources. This step will prevent inconsistencies stemming from incomplete data alignment and provide a more reliable basis for factor construction and portfolio analysis.

## Handling Missing Values

The cleaned dataset indicates missing values, particularly for Compustat variables. These may arise either from linking issues or incomplete company reporting. For CRSP data, forward-filling has been applied to address gaps, but missing Compustat values have been left unchanged to avoid introducing artificial data. Future iterations will explore imputation techniques based on industry or sector averages to ensure consistency without overestimating or underestimating firm performance.

## Resolving Ambiguous Permno-Gvkey Mappings

Approximately 3% of permnos (480 out of 16,417) exhibit multiple gvkeys, likely reflecting company restructurings, mergers, or acquisitions. To maintain processability, we opted to retain only those permnos representing companies listed in the S&P 500 during the analysis period, dropping the remainder. While effective for this study, this approach introduces a limitation by excluding potentially significant firms. Future improvements will focus on developing a more nuanced methodology to handle such cases, potentially assigning weights to gvkeys based on the duration or significance of their linkage to a permno.

# References

[1] Momentum: Studies like Jegadeesh and Titman (1993) document that winners tend to continue winning during bullish periods.

[2] Value: Fama and French's (1992) three-factor model highlights the outperformance of undervalued stocks over time.

[3] Size: Smaller firms tend to thrive in periods of economic growth as their risk premium decreases.

[4] Past Returns: Medium-term momentum and short-term reversals often perform well in optimistic markets.

[5] Quality Minus Junk (QMJ): Asness, Frazzini, and Pedersen (2013) demonstrate that quality firms with stable earnings outperform in downturns.

[6] Profitability: Novy-Marx (2013) shows that profitability is a critical driver of long-term equity returns, especially during recessions.

[7] Earnings Quality: Companies with low accruals and transparent earnings fare better in turbulent times, as highlighted by Sloan (1996).

[8] Distress Risk: Portfolios targeting low leverage and financially stable firms (e.g., Altman's Z-Score) reduce exposure to financial distress.

[9] Macro Sensitivity: Exposure to factors like interest rates, inflation, and VIX helps navigate macro-driven market movements.

[10] Liquidity: Amihud's (2002) Illiquidity Factor shows that liquid assets outperform during volatility due to higher trading costs for illiquid assets.

[11] Distress Risk: Maintaining exposure to low-risk, financially stable firms provides downside protection.

[12] Academic literature supports the diversification of factors to reduce unsystematic risk (Markowitz, 1952).