Multifactor Stock Analysis System (MSAS)

Complete System Documentation

Version 3.0 | September 2025

Table of Contents

PART I: MARKET REGIME DETECTION SYSTEM

- 1.1 Regime Detection System Overview
- 1.2 Core Algorithm
  - 1.2.1 Feature Engineering
  - 1.2.2 PCA Dimensionality Reduction
  - 1.2.3 Hidden Markov Models
- 1.3 Market Regime Types
- 1.4 Validation Framework
  - 1.4.1 Known Events Validation
  - 1.4.2 VIX Regime Comparison
- 1.5 Technical Implementation
- 1.6 System Integration

PART II: MULTIFACTOR STOCK SCORING SYSTEM

- 2.1 Multifactor Stock Scoring System Overview
- 2.2 Scoring Methodology
- 2.3 The 12 Factors - Detailed Analysis
  - 2.3.1 Value Factor
  - 2.3.2 Quality Factor
  - 2.3.3 Financial Health Factor
  - 2.3.4 Momentum Factor
  - 2.3.5 Growth Factor
  - 2.3.6 Technical Factor
  - 2.3.7 Stability Factor
  - 2.3.8 Size Factor
  - 2.3.9 Credit Factor
  - 2.3.10 Liquidity Factor
  - 2.3.11 Carry Factor

2.3.12 Insider Factor

2.4 Market Regime Weights

2.5 Operating Modes

2.6 Technical Implementation

PART III: SCORE TREND ANALYSIS SYSTEM

3.1 Score Trend Analysis System Overview

3.2 Core Algorithm: Stability-Adjusted Score

3.3 Regression Analysis Components

3.4 Composite Stability Score

3.5 Technical Indicator Integration

3.6 Technical Scoring System

3.7 Investment Recommendation Engine

3.8 Visualization and Reporting

3.9 Implementation Details

PART IV: DYNAMIC PORTFOLIO SELECTION SYSTEM

4.1 Dynamic Portfolio Selection System Overview

4.2 Core Selection Algorithm

4.3 Key Configuration Parameters

4.4 Portfolio Construction Strategies

4.5 Sector-Balanced Selection Algorithm

4.6 Hybrid Stability-Sector Portfolio

4.7 Industry and Sector Exclusions

4.8 Backtesting Engine

4.9 Risk and Performance Metrics

4.10 Portfolio Optimization Process

4.11 Advanced Features

4.12 Performance Characteristics

PART V: ADVANCED PORTFOLIO OPTIMIZATION SYSTEM

5.1 Advanced Portfolio Optimization System Overview

5.2 Theoretical Foundation: HRP

5.3 Mathematical Framework of HRP

5.4 Covariance Estimation: Ledoit-Wolf

5.5 HERC: Hierarchical Equal Risk Contribution

5.6 MHRP: Modified HRP

5.7 NCO: Nested Clustered Optimization

5.8 Method Comparison Guide

5.9 Implementation Architecture

5.10 Risk Metrics & Performance

5.11 Practical Considerations

5.12 Academic References

PART I: Market Regime Detection System

Hidden Markov Model-Based Market State Classification

1.1 Regime Detection System Overview

The Market Regime Detection System is a sophisticated quantitative framework that identifies and classifies market conditions using Hidden Markov Models (HMM). This system serves as the foundation for adaptive portfolio management, enabling regime-specific factor weighting and risk management strategies.

Key Capabilities:

- Real-time market regime identification across three distinct states
- 10-year historical regime analysis and characterization
- Validation accuracy of 80-90% against known market events
- Integration with multifactor scoring for adaptive weighting
- Probabilistic regime transition predictions

Data Collection → **Feature Engineering** → **PCA Reduction** → **HMM Training** → **Regime Classification**

1.2 Core Algorithm

1.2.1 Feature Engineering

The system constructs a comprehensive feature set capturing multiple dimensions of market behavior. These features are designed to capture regime-specific characteristics that distinguish between different market states.

Return Features

- **1-day returns:** Immediate price movements
- **5-day returns:** Weekly momentum
- **21-day returns:** Monthly trend

Rolling Return:

```
r_n = mean(returns[t:t-n])
```

Volatility Metrics

- **21-day volatility:** Short-term risk
- **63-day volatility:** Quarterly stability
- **Volatility change:** Risk regime transitions

Annualized Volatility:

```
σ = std(returns) × √252
```

Market Stress Indicators

- **VIX normalization:** Fear gauge z-score
- **Drawdown depth:** Peak-to-trough decline
- **Recovery time:** Drawdown duration

VIX Z-Score:

```
z = (VIX - μ_252) / σ_252
```

Trend Indicators

- **SMA crossover:** 50/200 day signal
- **Price momentum:** Rate of change
- **Trend strength:** ADX indicator

Trend Signal:

```
trend = (SMA_50 - SMA_200) / SMA_200
```

Cross-Asset Signals

- **Bond correlation:** Flight-to-quality
- **Gold correlation:** Safe haven demand
- **Dollar strength:** Risk-off indicator

Rolling Correlation:

```
ρ = corr(SPY, Asset)[60d]
```

Market Microstructure

- **Volume patterns:** Participation levels
- **Volatility term structure:** Forward-looking risk
- **Skewness:** Tail risk asymmetry

Volume Ratio:

```
v_ratio = Volume / SMA(Volume, 50)
```

1.2.2 PCA Dimensionality Reduction

Principal Component Analysis (PCA) is employed to reduce the dimensionality of the feature space while preserving the maximum amount of variance. This step is crucial for improving HMM convergence and preventing overfitting.

PCA Implementation Process:

1. **Standardization:** Features are standardized to zero mean and unit variance using

StandardScaler

2. **Component Selection:** Retain 5 principal components capturing ~85-90% of total variance
3. **Transformation:** Project original features onto principal component space
4. **Interpretation:** Analyze component loadings to understand regime drivers

Principal Component	Variance Explained	Cumulative Variance	Primary Loadings
PC1	42.3%	42.3%	Volatility, VIX, Drawdown
PC2	23.7%	66.0%	Trend, Momentum, Returns
PC3	12.4%	78.4%	Cross-asset correlations
PC4	7.8%	86.2%	Volume, Microstructure
PC5	5.2%	91.4%	Skewness, Tail risk

1.2.3 Hidden Markov Models

The HMM framework models market regimes as hidden states that generate observable market features through Gaussian emissions. The system uses a 3-state model optimized through extensive backtesting.

HMM Architecture:

- **Number of States:** 3 (Crisis/Bear, Steady Growth, Strong Bull)
- **Emission Type:** Multivariate Gaussian with full covariance
- **Training Algorithm:** Baum-Welch (Expectation-Maximization)
- **Initialization:** K-means clustering for robust starting points
- **Convergence Criteria:** Log-likelihood improvement < 0.01 or 200 iterations

Stability Enhancement: The system employs multiple random initializations (5 attempts with different seeds) to ensure convergence to the global optimum. The model with the highest log-likelihood is selected.

1.3 Market Regime Types

Crisis/Bear Regime

Frequency: 5-10% of historical periods

Characteristics:

- VIX > 30 (often > 40)
- Annualized volatility > 25%
- Negative returns: -20% to -40% annualized
- High correlation across assets
- Elevated drawdowns > 10%

Examples: COVID crash (2020), Volmageddon (2018), Financial Crisis (2008)

Steady Growth Regime

Frequency: 60-70% of historical periods

Characteristics:

- VIX: 12-20
- Moderate volatility: 10-18%
- Positive returns: 8-15% annualized
- Normal correlations
- Shallow drawdowns < 5%

Examples: 2023 recovery, 2016-2017 expansion, 2013-2014 growth

Strong Bull Regime

Frequency: 20-30% of historical periods

Characteristics:

- VIX < 12
- Low volatility: 8-12%
- Strong returns: 15-25% annualized
- Momentum persistence
- Minimal drawdowns < 3%

Examples: 2017 melt-up, 2021 bull run, 2019 Fed pivot rally

1.4 Validation Framework

1.4.1 Known Events Validation

The system validates regime detection accuracy against a comprehensive database of known market events, achieving 80-90% classification accuracy on historical data.

Historical Event Classification Results

Event	Period	Expected Regime	Detected Regime	Match
COVID Crash	2020-02-19 to 2020-03-23	Crisis/Bear	Crisis/Bear	✓
2017 Melt-Up	2017-01-01 to 2017-12-31	Strong Bull	Strong Bull	✓
2023 Recovery	2023-03-01 to 2023-12-31	Steady Growth	Steady Growth	✓

Overall Accuracy: 90.0% (9/10 correct classifications)

1.4.2 VIX Regime Comparison

The system's regime classifications are cross-validated against VIX-based regime definitions to ensure consistency with market-implied volatility expectations.

VIX Level	VIX Regime	Typical HMM Regime	Alignment Rate
< 12	Low Volatility	Strong Bull	92%
12-20	Normal	Steady Growth	85%
20-30	Elevated	Steady Growth / Transition	78%
> 30	High Volatility	Crisis/Bear	95%

Validation Success: The HMM-based regime detection shows strong alignment with both historical events (90% accuracy) and VIX-based classifications (87% average alignment), confirming the model's reliability for market regime identification.

1.5 Technical Implementation

Core Python Dependencies:

- **hmmlearn 0.3.0:** Hidden Markov Model implementation
- **scikit-learn 1.3.0:** PCA and preprocessing
- **pandas 2.0.3:** Data manipulation and time series
- **numpy 1.24.3:** Numerical computations
- **yfinance 0.2.28:** Market data retrieval
- **matplotlib 3.7.2:** Visualization
- **PyQt5 5.15.9:** GUI framework

File Structure:

- `regime_detector.py` - Core HMM implementation
- `current_regime_detector.py` - Real-time detection
- `regime_detection_widget.py` - PyQt5 GUI component
- `output/regime_periods.csv` - Historical regime classifications
- `output/regime_model.pkl` - Trained HMM model

1.6 System Integration

The Market Regime Detection system seamlessly integrates with other MSAS components to provide adaptive, regime-aware investment strategies.

Integration Points:

1. Multifactor Scoring System:

- Regime detection triggers automatic weight adjustment
- Crisis/Bear: Defensive factors (Quality, Size) emphasized
- Strong Bull: Growth and Momentum factors prioritized
- Steady Growth: Balanced factor weighting

2. Portfolio Optimization:

- Risk parameters adjusted based on detected regime
- Position sizing modified for regime-specific volatility
- Rebalancing frequency adapted to market conditions

3. Score Trend Analysis:

- Technical indicators calibrated for regime volatility
- Lookback periods adjusted for regime characteristics
- Signal thresholds modified based on market state

PART II: Multifactor Stock Scoring System

Adaptive Factor-Based Security Ranking and Selection

2.1 Multifactor Stock Scoring System Overview

The Multifactor Scoring Analysis System (MSAS) is a sophisticated quantitative stock screening and ranking system that evaluates securities across 12 distinct factors, each composed of multiple subfactors. The system employs regime-specific weighting schemes optimized through historical backtesting to adapt to varying market conditions.

Key Features:

- 12 primary factors with 50+ subfactors
- Continuous scoring (0-1 scale) for all metrics
- Industry and sector-relative adjustments
- Three market regime configurations
- Multiple operational modes for current and historical analysis

2.2 Scoring Methodology

Score Calculation Pipeline

The scoring process follows a sophisticated multi-stage pipeline:

1. **Data Collection:** Retrieve fundamental, technical, and alternative data from multiple sources (Yahoo Finance, Marketstack, SEC filings)
2. **Factor Calculation:** Compute individual factor scores using continuous functions
3. **Normalization:** Apply cross-sectional normalization and sector/industry adjustments
4. **Weighting:** Apply regime-specific weights to calculate composite scores
5. **Ranking:** Sort securities by final composite score

Composite Score:

```
Score = Σ(Factor_i × Weight_i) for i = 1 to 12
```

Continuous Scoring Functions

All factors utilize continuous scoring functions that map raw metrics to a [0,1] range, ensuring smooth transitions and avoiding binary thresholds:

General Scoring Function:

```
Score = max(0, min(1, (Value - Poor) / (Excellent - Poor)))
```

Cross-Sectional Normalization

Momentum and volatility metrics undergo z-score normalization followed by sigmoid transformation:

Z-Score Normalization:

```
z = (x - μ) / σ
```

Sigmoid Transform:

```
score = 1 / (1 + exp(-z))
```

2.3 The 12 Factors - Detailed Analysis

2.3.1 Value Factor Variable (-18.7 to -7.3)

Purpose: Identifies undervalued securities relative to fundamentals. Uses industry-relative adjustments to account for sector-specific valuation norms.

Earnings Yield (Weight: 25%)

Inverse of P/E ratio, measuring earnings return on investment

```
Earnings Yield = 1 / P/E Ratio
```

Threshold: P/E < 20 (excellent), > 30 (poor)

Free Cash Flow Yield (Weight: 25%)

Cash generation relative to market capitalization

```
FCF Yield = Free Cash Flow / Market Cap
```

Threshold: > 5% (excellent), < 2% (poor)

EV/EBITDA (Weight: 20%)

Enterprise value relative to operating earnings

```
Score = 1 - tanh((EV/EBITDA - 12) / 8)
```

Threshold: < 8 (excellent), > 16 (poor)

Price-to-Book (Weight: 15%)

Market value relative to book value

```
Score = 1 - tanh((P/B - 2.5) / 2)
```

Threshold: < 1.5 (excellent), > 4 (poor)

EV/Sales (Weight: 15%)

Enterprise value relative to revenue

Threshold: < 1 (excellent), > 3 (poor)

Industry Adjustment: All value metrics undergo industry-relative z-score normalization to account for sector-specific valuation differences.

2.3.2 Quality Factor Variable (-13.3 to 18.1)

Purpose: Measures operational excellence and profitability metrics.

Return on Equity (ROE) (Weight: 20%)

Profitability relative to shareholder equity

```
Score = (ROE - 6%) / 24%
```

Threshold: > 30% (excellent), < 6% (poor)

Return on Invested Capital (ROIC) (Weight: 20%)

Efficiency of capital deployment

```
ROIC = EBIT / (Total Assets - Current Liabilities)
```

Threshold: > 20% (excellent), < 5% (poor)

Gross Margin (Weight: 15%)

Pricing power and cost efficiency

Threshold: > 50% (excellent), < 20% (poor)

FCF Margin (Weight: 15%)

Cash generation efficiency

```
FCF Margin = Free Cash Flow / Revenue
```

Threshold: > 15% (excellent), < 3% (poor)

Asset Turnover (Weight: 10%)

Revenue generation efficiency

```
Asset Turnover = Revenue / Total Assets
```

Threshold: > 1.5 (excellent), < 0.3 (poor)

Operating Margin (Weight: 10%)

Operational efficiency

Threshold: > 20% (excellent), < 5% (poor)

Revenue Stability (Weight: 10%)

Consistency of revenue generation

Calculated using 3-year revenue volatility

2.3.3 Financial Health Factor 14.4 to 64.0

Purpose: Assesses balance sheet strength and solvency.

Cash/Debt Ratio (Weight: 25%)

Liquidity relative to debt obligations

```
Cash/Debt = Total Cash / Total Debt
```

Threshold: > 60% (excellent), < 10% (poor)

Current Ratio (Weight: 25%)

Short-term liquidity

```
Current Ratio = Current Assets / Current Liabilities
```

Threshold: > 2.5 (excellent), < 1.0 (poor)

Debt-to-Equity (Weight: 25%)

Leverage assessment (lower is better)

Threshold: < 50 (excellent), > 150 (poor)

Interest Coverage (Weight: 25%)

Ability to service debt

```
Interest Coverage = EBIT / Interest Expense
```

Threshold: > 10 (excellent), < 2 (poor)

2.3.4 Momentum Factor 18.1 to 72.0

Purpose: Captures price and earnings momentum trends.

Price Momentum (Weight: 60%)

252-day (12-month minus 1-month) price return

```
PM = (Price_current / Price_252d_ago) - 1
```

Cross-sectional z-score normalization applied

Earnings Momentum (Weight: 40%)

Combination of estimate revisions and earnings surprises

```
EM = 0.5 × Revision% + 0.5 × Surprise%
```

90-day estimate changes and trailing 4-quarter surprises

Normalization: Both components undergo z-score normalization followed by sigmoid transformation for final score.

2.3.5 Growth Factor 5.0 to 15.0

Purpose: Forward-looking growth assessment with sector adjustments.

Revenue Growth (Weight: 0.8)

Year-over-year revenue growth rate

Threshold: > 30% (excellent), < 5% (poor)

PEG Ratio (Weight: 0.9)

Price/Earnings relative to growth (inverted)

```
Score = max(0, min(1, (2 - PEG) / 1.5))
```

Threshold: < 1 (excellent), > 2 (poor)

Earnings Growth Estimate (Weight: 1.0)

Analyst consensus forward growth

Based on next 12 months estimates

Price Target Upside (Weight: 0.7)

Analyst price target relative to current price

```
Upside = (Target Price / Current Price) - 1
```

Operating Leverage (Weight: 0.6)

Margin expansion potential

Measured by operating margin trends

Sector Adjustment: Growth scores undergo sector-relative z-score normalization to account for industry-specific growth expectations.

2.3.6 Technical Factor -5.0 to 8.0

Purpose: Technical analysis indicators for trend and momentum.

Breakout Score (Weight: 35%)

52-week high proximity and volume confirmation

```
Breakout = Price_proximity × Volume_ratio
```

Trend Score (Weight: 40%)

Multiple trend indicators combined

- ADX strength (> 25 for strong trend)
- MACD signal alignment
- RSI momentum (30-70 optimal range)

Moving Average Score (Weight: 25%)

Price position relative to key moving averages

- 50-day SMA position
- 200-day SMA position
- Golden/Death cross signals

2.3.7 Stability Factor -10.0 to -3.0

Purpose: Risk assessment through volatility and drawdown metrics. Higher stability = lower risk = better score.

Volatility Score (Weight: 30%)

Annualized standard deviation of returns

```
Vol Score = 1 - tanh(σ_annual × 2)
```

Lower volatility results in higher score

Tail Risk Score (Weight: 30%)

Downside risk measures

- Conditional Value at Risk (CVaR 95%)
- Conditional Drawdown at Risk (CDaR)

Recovery Speed (Weight: 20%)

Speed of recovery from drawdowns

Weighted average recovery time from 5%+ drawdowns

Beta Score (Weight: 20%)

Market sensitivity

```
Beta Score = 1 - tanh(max(0, β - 1) × 2)
```

Lower beta (< 1) preferred

2.3.8 Size Factor -2.2 to 20.0

Purpose: Size premium capture with liquidity considerations.

Market Cap Score (Weight: 80%)

Preference for smaller capitalization

```
MC Score = 1 - (log(MC) - log(MC_min)) / (log(MC_80th) - log(MC_min))
```

Uses 80th percentile as upper threshold

Price Level Score (Weight: 20%)

Optimal price range assessment

Penalizes extreme prices (< $2 or > $500)

Optimal range: $10-$100

2.3.9 Credit Factor 0.0 to 48.6

Purpose: Credit quality and default risk assessment.

Altman Z-Score Components

Bankruptcy prediction model

```
Z = 1.2×WC/TA + 1.4×RE/TA + 3.3×EBIT/TA + 0.6×MC/TL + 1.0×S/TA
```

- Working Capital / Total Assets
- Retained Earnings / Total Assets
- EBIT / Total Assets
- Market Cap / Total Liabilities
- Sales / Total Assets

Score > 3.0 (safe), 1.8-3.0 (gray zone), < 1.8 (distress)

2.3.10 Liquidity Factor 1.9 to 5.0

Purpose: Trading liquidity assessment.

Volume Score

10-day average volume relative to peers

```
Liquidity = (Vol - Vol_min) / (Vol_max - Vol_min)
```

Cross-sectional normalization across universe

2.3.11 Carry Factor -5.0 to -0.9

Purpose: Dividend yield and sustainability assessment.

Dividend Yield (Weight: 60%)

Current dividend return

```
Base Score = 0.25 + min(max((Yield - 3%) / 5%, 0), 0.75)
```

Threshold: 3-8% optimal range

Payout Ratio (Weight: 20%)

Dividend sustainability

Threshold: < 70% (sustainable), > 100% (unsustainable)

Dividend Growth (Weight: 20%)

5-year dividend growth rate

Positive growth adds to score

2.3.12 Insider Factor -2.0 to 5.0

Purpose: Insider trading sentiment analysis.

Net Insider Activity

90-day insider transaction analysis

```
Score = (Buys - Sells) / (Buys + Sells + 1)
```

- Purchase transactions: +1
- Sale transactions: -1
- Weighted by transaction value

Unique Insider Count

Number of distinct insiders transacting

Higher participation indicates stronger signal

Transaction Recency

Time-decay weighting for transactions

More recent transactions receive higher weight

2.4 Market Regime Weights

The system employs three distinct weight configurations optimized for different market regimes. These weights are derived from extensive backtesting and correlation analysis across historical periods.

Steady Growth Regime

Credit **48.6**

Quality **18.1**

Momentum **18.1**

Financial Health **14.4**

Growth **10.9**

Technical **3.8**

Liquidity **3.6**

Insider **-1.4**

Size **-2.2**

Stability **-3.0**

Carry **-3.6**

Value **-7.3**

Strong Bull Regime

Credit **40.0**

Momentum **25.0**

Quality **20.0**

Financial Health **15.0**

Growth **15.0**

Technical **8.0**

Liquidity **5.0**

Insider **2.0**

Size **0.0**

Carry **-2.0**

Value **-5.0**

Stability **-5.0**

Crisis/Bear Regime

Momentum **72.0**

Size **20.0**

Financial Health **13.0**

Insider **5.0**

Growth **5.0**

Liquidity **5.0**

Credit **0.0**

Quality **0.0**

Value **0.0**

Technical **-5.0**

Carry **-5.0**

Stability **-10.0**

Weight Interpretation:

- Positive weights indicate factors that contribute positively to expected returns in that regime
- Negative weights represent contrarian factors or risk penalties
- Zero weights indicate factors with no predictive power in that regime
- Weights are normalized to sum to approximately 100 for interpretability

2.5 Operating Modes

The system supports multiple operational modes to accommodate different analysis requirements:

Available Modes:

1. Daily Current Mode (Default)

- Real-time analysis using current market data
- Automatic regime detection integration
- Output: top_ranked_stocks_[regime]_[date].csv

2. Historical Backtesting Mode

- Point-in-time analysis for specific dates
- Used for strategy validation and optimization
- Supports date range: 2015-present

3. Batch Processing Mode

- Multiple date analysis for trend identification
- Generates time series of factor scores
- Used for stability analysis and factor research

4. Research Mode

- Full factor decomposition output
- Individual subfactor scores exported
- Used for factor development and optimization

2.6 Technical Implementation

System Architecture:

- **Data Layer:** Yahoo Finance, Marketstack, SEC EDGAR integration
- **Calculation Engine:** Vectorized NumPy/Pandas operations
- **Caching System:** Redis-based for API call optimization
- **Output Format:** CSV with full factor decomposition
- **Performance:** ~900-1000 stocks processed in 5-10 minutes

Output File Structure:

- Ticker: Stock symbol
- CompanyName: Full company name
- Country: Headquarters location
- Score: Final composite score
- Value, Quality, FinancialHealth, etc.: Individual factor scores
- Sector: GICS sector classification
- Industry: GICS industry classification

Important Notes:

- Factor weights are based on extensive backtesting across multiple market cycles
- Negative weights indicate contrarian factors in specific regimes
- The system processes approximately 900-1000 stocks per run
- Industry and sector adjustments are critical for value and growth factors
- All scores are normalized to [0,1] range for comparability

PART III: Score Trend Analysis System

Stability Assessment and Technical Analysis for Stock Score Trends

3.1 Score Trend Analysis System Overview

The Score Trend Analysis System provides comprehensive stability assessment and technical analysis for stocks that have appeared in multiple ranking periods. It combines regression analysis, stability metrics, and technical indicators to identify stocks with consistent performance and positive momentum.

Key Capabilities:

- Time series analysis of factor scores across multiple periods
- Linear and polynomial regression for trend identification
- Stability-adjusted scoring with multiple adjustment layers
- 14 technical indicators across 5 categories
- Investment recommendations based on composite metrics
- Visual analysis with multi-panel technical charts

3.2 Core Algorithm: Stability-Adjusted Score

The system employs a sophisticated multi-stage adjustment process to calculate the final stability-adjusted score, which represents both the quality and consistency of a stock's performance.

Score Adjustment Pipeline:

The final stability-adjusted score is calculated through four progressive adjustments:

1. **Base Score:** Average score across all appearances
2. **R^2 Adjustment:** Quality of linear trend
3. **Slope Adjustment:** Direction and strength of trend
4. **Stability Adjustment:** Consistency and volatility penalty

Step 1: Average Score Calculation

The foundation metric representing overall quality:

Average Score:

```
avg_score = Σ(scores) / n_appearances
```

Range: 0-100

Purpose: Baseline quality measure

Step 2: R^2 Adjusted Score

Adjusts for trend reliability using coefficient of determination:

R² Adjusted Score:

```
r2_adjusted = avg_score × linear_r2
```

Linear R^2: Measures how well a linear model fits the data

Impact: Penalizes erratic score patterns

Step 3: Slope Adjusted Score

Incorporates trend direction using sigmoid weighting:

Sigmoid Factor:

```
sigmoid = 1 / (1 + e^(-k × slope))
```

Slope Adjusted:

```
slope_adjusted = r2_adjusted × (0.5 + sigmoid)
```

k (sensitivity): Default 5.0

Range: 0.5× to 1.5× multiplier

Step 4: Stability Adjusted Score

Final adjustment for consistency using standard deviation:

Stability Factor:

```
stability = 1 / (1 + score_std)
```

Final Score:

```
stability_adjusted = slope_adjusted × stability
```

Purpose: Penalizes volatile performers

Result: Risk-adjusted performance metric

3.3 Regression Analysis Components

Linear Regression Analysis:

The system performs linear regression on the time series of scores to identify trends:

Linear Model:

```
Score = β  + β  × Time_Index + ε
```

Key Metrics Extracted:

- **Slope (β_1):** Rate of score change per period
- **Intercept (β_0):** Initial score estimate
- **R^2 Score:** Goodness of fit (0-1 scale)
- **Predictions:** Fitted values for visualization

Metric	Calculation	Interpretation	Ideal Range
Linear Slope	Δscore/Δtime	Trend direction and strength	> 0.5 (strong uptrend)
R^2 Score	1 - (SS_res/SS_tot)	Trend consistency	> 0.7 (high reliability)
Score Std Dev	σ(scores)	Volatility measure	< 5.0 (low volatility)
Coefficient of Variation	σ/μ	Normalized volatility	< 0.1 (stable)
Trend Consistency	% periods with positive Δ	Directional consistency	> 0.7 (consistent growth)

3.4 Composite Stability Score

In addition to the stability-adjusted score, the system calculates a composite stability score (0-100) that provides a holistic assessment of reliability:

Composite Stability Score Components:

```
Stability Score =
R²_score + CV_score + Consistency_score + Slope_bonus
```

Component Breakdown:

- **R^2 Score:** linear_r2 × 30 (max 30 points)
- **CV Score:** max(0, 30 - CV × 100) (max 30 points)
- **Consistency Score:** trend_consistency × 20 (max 20 points)
- **Slope Bonus:** min(20, slope × 1000) if slope > 0 and R^2 > 0.5 (max 20 points)

Total Range: 0-100 points (capped)

3.5 Technical Indicator Integration

The system calculates 14 technical indicators grouped into 5 categories, each contributing to a comprehensive technical score:

Trend Analysis Indicators

- **Trend Direction:** Linear regression slope classification
- **Trend Quality:** R^2 score thresholds
- **ADX Strength:** Average Directional Index (> 25 = strong)
- **DI Direction:** +DI/-DI crossover signals

Moving Average Indicators

- **SMA Position:** Price vs 5/10/20-day SMAs
- **Recent Crossover:** Bullish/bearish MA crossovers
- **EMA Signals:** Exponential moving average trends

Momentum Indicators

- **MACD Signal:** MACD vs signal line position
- **MACD Histogram:** Momentum acceleration
- **RSI Level:** Overbought/oversold conditions
- **RSI Trend:** RSI directional movement
- **Rate of Change:** Momentum velocity

Volatility Indicators

- **Bollinger %B:** Position within bands
- **Band Width:** Volatility expansion/contraction
- **ATR:** Average True Range for volatility

Forecasting Models

- **ARIMA Forecast:** Time series prediction
- **Exponential Smoothing:** Trend projection

3.6 Technical Scoring System

Technical Score Calculation:

Each indicator contributes 0-1 points based on signal strength:

Indicator	Bullish Signal (+1)	Neutral Signal (+0.5)	Bearish Signal (0)
Trend Direction	Slope > 0.5	0 < Slope < 0.5	Slope < 0
Trend Quality	$R^2 > 0.7$	$0.4 < R^2 < 0.7$	$R^2 < 0.4$
RSI Level	60 < RSI < 70	40 < RSI < 60	RSI < 30 or > 70
MACD Signal	MACD > Signal	Convergence zone	MACD < Signal
MA Position	Price > SMA5 > SMA10	Mixed alignment	Price < both MAs
ADX Strength	ADX > 25 with +DI > -DI	20 < ADX < 25	ADX < 20 or -DI > +DI

Technical Score:

```
Score = (Σ indicator_points / max_possible) × 100
```

3.7 Investment Recommendation Engine

The system generates investment recommendations based on the combination of stability-adjusted scores and technical indicators:

Recommendation Thresholds:

Recommendation	Stability Adjusted Score	Additional Criteria
STRONG BUY - Elite	≥ 40	Slope > 0.3 AND $R^2 > 0.7$
STRONG BUY	≥ 35	Slope > 0.2 AND $R^2 > 0.6$
BUY	≥ 30	Slope > 0.1 OR Avg Score > 60
HOLD	≥ 25	|Slope| < 0.1
WATCH	20-25	Monitor for opportunity
REDUCE	< 20	Slope < -0.2
SELL	< 15	Slope < -0.3 AND $R^2 > 0.6$

Special Conditions:

- **Volatility Override:** "AVOID" if CV > 0.15 (too volatile)
- **Momentum Play:** "SPECULATIVE BUY" if slope > 0.5 despite lower stability
- **Quality Hold:** "HOLD - Quality" if avg_score > 70 but flat trend

3.8 Visualization and Reporting

The system generates comprehensive visual reports including:

Multi-Panel Technical Charts

- Score trend with regression lines
- Moving averages with crossover signals
- RSI with overbought/oversold zones
- MACD with signal and histogram
- Bollinger Bands with %B indicator
- ADX trend strength indicator

Output Reports

- CSV rankings with all metrics
- Individual stock technical reports
- Sector/industry analysis
- Top stable performers summary
- High volatility warnings
- Trend comparison matrices

3.9 Implementation Details

System Configuration:

Data Requirements:

- Minimum 3 appearances for analysis
- Ideal: 10+ periods for reliable trends
- Input: Historical ranking CSV files
- Lookback: Configurable (default 90 days)

Technical Stack:

- **NumPy:** Numerical computations
- **Pandas:** Data manipulation
- **Scikit-learn:** Regression analysis
- **Matplotlib:** Visualization
- **Statsmodels:** ARIMA forecasting

Performance Metrics:

- Processing time: ~1-2 seconds per stock
- Typical analysis: 100-200 stocks
- Output files: 5-10 MB typical

Key Innovation: The multi-layered adjustment approach (R^2 → Slope → Stability) provides a sophisticated risk-adjusted performance metric that balances quality, trend, and consistency - crucial for identifying stocks with sustainable competitive advantages.

Usage Notes:

- Best suited for stocks with consistent ranking appearances
- Technical indicators require sufficient data points (minimum 20)
- Recommendations should be combined with fundamental analysis
- Stability scores are relative within the analyzed universe

PART IV: Dynamic Portfolio Selection System

Intelligent Portfolio Construction from Stability-Ranked Securities

4.1 Dynamic Portfolio Selection System Overview

The Dynamic Portfolio Selection System creates optimized portfolios from stability-analyzed stocks using multiple construction strategies, diversification constraints, and sophisticated backtesting algorithms. The system adapts its selection methodology based on market regime and employs various optimization criteria to balance risk, return, and diversification.

Key Capabilities:

- Multiple portfolio construction strategies (9+ methods)
- Configurable pool sizes and portfolio dimensions
- Sector and industry diversification constraints
- Multi-period backtesting with rebalancing strategies
- Risk-adjusted performance optimization
- Real estate and industry-specific exclusions

4.2 Core Selection Algorithm

The system employs a hierarchical selection process that filters and ranks stocks through multiple stages:

Stability Analysis Results → **Top Pool Selection** → **Strategy Application** → **Constraint Enforcement** → **Portfolio Construction**

4.3 Key Configuration Parameters

Top Pools

Definition: Number of top-ranked stocks to consider as selection universe

Common Values: [50, 75, 100, 150]

Purpose: Controls quality threshold

Pool Selection:

```
Pool = Top N stocks by stability_adjusted_score
```

Portfolio Sizes

Definition: Target number of stocks in final portfolio

Common Values: [10, 15, 20, 25]

Purpose: Balance diversification vs concentration

Size Constraint:

```
10 ≤ Portfolio_Size ≤ min(Pool_Size, 30)
```

Max Stocks Per Sector

Definition: Maximum concentration allowed per sector

Common Values: 2-3 stocks

Purpose: Enforce sector diversification

Sector Constraint:

```
Count(Sector_i) ≤ Max_Per_Sector
```

Lookback Periods

Definition: Historical days for backtesting

Common Values: [252, 504, 756] days

Purpose: Performance validation window

Lookback Window:

```
T_start = T_current - Lookback_Days
```

4.4 Portfolio Construction Strategies

The system implements nine distinct portfolio construction strategies, each optimized for different objectives:

Strategy Implementation Matrix:

Strategy	Selection Criteria	Optimization Goal	Best For
TopRank	Highest stability_adjusted_score	Quality maximization	Core holdings
SectorBal2/3	Top scores with max 2-3 per sector	Sector diversification	Risk reduction
IndustryBal	Top scores with max 2 per industry	Industry diversification	Granular diversification
Stability	Highest R^2 values (≥ 0.5)	Consistency maximization	Low volatility

PosMomentum	Positive linear_slope only	Growth capture	Trending markets
Recommended	BUY recommendations only	Signal strength	High conviction
HybridStability	Weighted R^2 and rank composite	Balanced optimization	All-weather
FactorBal	Balanced factor exposures	Factor diversification	Multi-factor
Diversified	Top 5 sectors, 4 stocks each	Maximum diversification	Conservative

4.5 Sector-Balanced Selection Algorithm

The sector-balanced strategy ensures diversification while maintaining quality through a constraint-based selection process:

Algorithm Logic:

1. **Initialize:**
    - Create empty portfolio list
    - Initialize sector counter dictionary
    - Set target portfolio size and max stocks per sector
2. **Sort Candidates:**
    - Order all stocks by stability_adjusted_score (descending)
    - This ensures highest quality stocks are considered first
3. **Iterative Selection:**
    - For each stock in sorted order:
        - → → Check if portfolio size reached (exit if true)
        - → → Get stock's sector classification
        - → → Check sector count against max_per_sector limit
        - → → If under limit: add stock and increment sector count
        - → → Otherwise: skip to next stock
4. **Termination:**
    - Stop when target portfolio size is reached

- Or when candidate pool is exhausted

Constraint Function:

```
Accept(stock) = (Portfolio_Size < Target) AND
(Sector_Count[stock.sector] < Max_Per_Sector)
```

Key Properties:

- **Complexity:** O(n) where n = pool size
- **Optimality:** Greedy selection ensures best available stocks within constraints
- **Guarantee:** Hard limit on sector concentration is always enforced
- **Flexibility:** Works with any scoring metric (stability_adjusted_score, R^2, etc.)

> **Example:** With portfolio_size=20 and max_per_sector=3, the algorithm might select stocks from 7-10 different sectors, ensuring no single sector dominates the portfolio while maintaining high average quality scores.

4.6 Hybrid Stability-Sector Portfolio

The most sophisticated strategy combines stability metrics with diversification constraints:

Hybrid Composite Score Calculation:

Step 1 - Normalize R^2:

```
R²_norm = (R² - R²_min) / (R²_max - R²_min)
```

Step 2 - Normalize Rank (inverted):

```
Rank_norm = 1 - ((Rank - Rank_min) / (Rank_max - Rank_min))
```

Step 3 - Composite Score:

```
Score = w × R²_norm + (1-w) × Rank_norm

where w = stability_weight (default: 0.6)
```

This approach balances trend reliability (R^2) with overall quality (rank), then applies sector constraints.

4.7 Industry and Sector Exclusions

The system provides sophisticated filtering mechanisms for risk management:

Real Estate Exclusion

Identified by:

- Sector: "Real Estate"
- Industry: Contains "REIT"
- Keywords: "property", "mortgage", "homebuilding"

Purpose: Avoid interest rate sensitive sectors

Custom Industry Exclusions

Common Exclusions:

- Airlines (cyclical risk)
- Biotechnology (binary events)
- Energy (commodity dependence)
- Regional Banks (regulatory risk)

Implementation: Pre-filtering before portfolio construction

4.8 Backtesting Engine

The system employs a sophisticated backtesting framework with multiple rebalancing strategies:

Backtesting Architecture:

1. Data Preparation:

- Fetch historical prices for lookback period
- Handle delisted tickers and missing data
- Align time series across all assets

2. Rebalancing Strategies:

Strategy	Trigger	Frequency	Transaction Costs
Monthly	Calendar-based	Every 21 trading days	~0.1% per rebalance

Drift	Weight deviation > 5%	Variable	Lower (fewer rebalances)
Volatility	Vol regime change	Event-driven	Adaptive

3. Performance Calculation:

```
Portfolio Return:

R_p(t) = Σ w_i(t) × R_i(t) - Costs(t)

where Costs(t) = fees × I_rebalance(t) + slippage
```

4.9 Risk and Performance Metrics

The system calculates comprehensive risk-adjusted performance metrics:

Return Metrics

- **Total Return:** (Final/Initial - 1)
- **Annual Return:** CAGR calculation
- **Monthly Returns:** Time series analysis
- **Win Rate:** % positive periods

Risk Metrics

- **Volatility:** Annualized std deviation
- **Max Drawdown:** Peak-to-trough
- **Drawdown Duration:** Recovery time
- **Beta:** Market sensitivity

Risk-Adjusted Metrics

- **Sharpe Ratio:** (R - Rf) / σ
- **Sortino Ratio:** Downside focus
- **Calmar Ratio:** Return / Max DD
- **Information Ratio:** Active return / TE

Diversification Metrics

- **Sector Count:** Unique sectors
- **Industry Count:** Unique industries
- **Concentration:** HHI index
- **Correlation:** Average pairwise

4.10 Portfolio Optimization Process

The complete portfolio selection and optimization workflow:

Selection Pipeline:

1. **Input Processing:**
   - Load stability analysis results (CSV)
   - Parse configuration parameters
   - Apply exclusion filters
2. **Portfolio Generation:**
   - For each (pool_size, portfolio_size) combination:
   - Generate all 9 strategy variants
   - Total portfolios = |pools| × |sizes| × |strategies|
3. **Backtesting:**
   - Fetch historical data (YFinance/Marketstack)
   - Apply rebalancing strategy
   - Calculate performance metrics
4. **Ranking and Selection:**
   - Sort by primary metric (Sharpe ratio)
   - Apply secondary filters (min return, max DD)
   - Select top N portfolios
5. **Output Generation:**
   - Performance comparison matrix
   - Risk-return scatter plots
   - Portfolio composition details
   - JSON export for further analysis

4.11 Advanced Features

Walk-Forward Analysis:

The system supports walk-forward optimization for out-of-sample validation:

```
Walk-Forward Window:

In-Sample: [T-lookback, T-forward]
Out-of-Sample: [T-forward, T]

Sharpe Degradation:

Degradation = (Sharpe_in - Sharpe_out) / Sharpe_in
```

Purpose: Detect overfitting and validate strategy robustness

4.12 Performance Characteristics

System Performance:

Operation	Time Complexity	Typical Duration
Portfolio Generation	O(pools × sizes × n)	< 1 second
Single Backtest	O(T × n)	2-5 seconds
Complete Analysis (100 portfolios)	O(P × T × n)	3-5 minutes
Walk-Forward (3-month windows)	O(W × P × T × n)	10-15 minutes

where: n = stocks, T = time periods, P = portfolios, W = windows

Key Innovation: The multi-strategy approach with configurable constraints allows the system to adapt to different market conditions and investor preferences while maintaining robust risk management through diversification requirements and comprehensive backtesting.

Best Practices:

- Use multiple pool sizes to test quality thresholds
- Maintain sector limits at 2-3 stocks for true diversification

- Prefer drift-based rebalancing to reduce transaction costs
- Validate with at least 2 years of historical data
- Consider regime-specific portfolio construction
- Export results for post-hoc analysis and reporting

PART V: Advanced Portfolio Optimization System

Hierarchical Risk Parity and Advanced Allocation Methods

5.1 Advanced Portfolio Optimization System Overview

The Portfolio Optimization System implements state-of-the-art hierarchical portfolio allocation methods that address the fundamental limitations of traditional mean-variance optimization. Building upon the pioneering work of Marcos López de Prado and subsequent researchers, this system provides robust, out-of-sample portfolio allocations through machine learning-inspired clustering techniques.

Key Methods Implemented:

- **HRP:** Hierarchical Risk Parity (López de Prado, 2016)
- **HERC:** Hierarchical Equal Risk Contribution (Raffinot, 2017)
- **MHRP:** Modified HRP with Equal Volatility Weighting
- **NCO:** Nested Clustered Optimization (López de Prado, 2020)

5.2 Theoretical Foundation: Hierarchical Risk Parity (HRP)

HRP revolutionized portfolio optimization by replacing the unstable matrix inversion of traditional methods with a hierarchical clustering approach inspired by machine learning. The method addresses three critical problems of mean-variance optimization: instability, concentration, and underperformance out-of-sample.

The HRP Algorithm - Three Stage Process:

Stage 1: Tree Clustering

- Transform correlation matrix into a distance matrix
- Apply hierarchical clustering to group similar assets
- Build a dendrogram representing asset relationships

Stage 2: Quasi-Diagonalization

- Reorder the covariance matrix to place similar assets together
- Creates a quasi-diagonal structure
- Preserves the hierarchical information from clustering

Stage 3: Recursive Bisection

- Allocate weights through top-down recursive splitting
- At each split, allocate between clusters based on inverse variance
- Continue until individual asset weights are determined

5.3 Mathematical Framework of HRP

Distance Matrix Calculation

Convert correlation to distance metric:

Correlation Distance:

$$d_{ij} = \sqrt{0.5 \times (1 - \rho_{ij})}$$

This ensures that perfectly correlated assets ($\rho=1$) have distance 0, while perfectly anti-correlated assets ($\rho=-1$) have distance 1.

Hierarchical Clustering

Ward's linkage method minimizes within-cluster variance:

Ward Distance:

$$d(u,v) = \sqrt{2 \times |u| \times |v| / (|u| + |v|) \times ||c_u - c_v||^2}$$

where c_u and c_v are cluster centroids

Quasi-Diagonalization

Recursive algorithm to order leaves of dendrogram:

1. Start with root cluster pair
2. Recursively expand clusters
3. Maintain order that preserves hierarchy

Result: Similar assets are adjacent in matrix

Recursive Bisection Allocation

Weight allocation based on inverse variance:

Allocation Ratio:

$$\alpha = 1 - \sigma^2_L / (\sigma^2_L + \sigma^2_R)$$

Left cluster gets weight α, right gets $(1-\alpha)$

5.4 Covariance Estimation: Ledoit-Wolf Shrinkage

All methods employ Ledoit-Wolf shrinkage to improve covariance matrix estimation, particularly crucial for high-dimensional portfolios or limited historical data:

Shrinkage Estimator:

Shrunk Covariance:

$$\Sigma^* = \delta F + (1-\delta) S$$

where:

- **S:** Sample covariance matrix (empirical)
- **F:** Shrinkage target (often diagonal with average variance)
- **δ:** Optimal shrinkage intensity ($0 \leq \delta \leq 1$)

Benefits:

- Reduces estimation error, especially with limited data
- Improves condition number of covariance matrix
- Better out-of-sample performance
- More stable weight allocations

5.5 HERC: Hierarchical Equal Risk Contribution

HERC enhances HRP by enforcing equal risk contribution at each hierarchical level, providing better diversification:

Key Differences from HRP:

Aspect	HRP	HERC
Risk Allocation	Inverse variance bisection	Equal risk contribution
Cluster Treatment	Simple splitting	Risk parity across clusters
Within-Cluster	Minimum variance	Equal risk contribution
Diversification	Good	Better (enforced equality)
Tail Risk	Moderate control	Superior control

Mathematical Foundation:

Risk Contribution:

$$RC_i = w_i \times (\Sigma w)_i / \sigma_p$$

HERC Constraint:

$$RC_i = RC_j \text{ for all } i, j \text{ in same cluster}$$

5.6 MHRP: Modified HRP with Equal Volatility

MHRP modifies the original HRP by using equal volatility weighting instead of inverse variance, providing more balanced allocations:

Core Modification

Original HRP: Weights ∝ $1/\sigma^2$

MHRP: Weights ∝ $1/\sigma$

MHRP Weight:

```
w_i = (1/σ_i) / Σ(1/σ_j)
```

Results in less extreme weight differences

Advantages

- More balanced allocations
- Reduced concentration in low-volatility assets
- Better performance in volatile markets
- Lower portfolio turnover
- More intuitive risk distribution

Cluster Volatility

For recursive bisection:

Cluster Vol:

```
σ_c = √(w'_c Σ_c w_c)
```

where w_c uses equal volatility weights within cluster

Robustness Features

- Spearman correlation for outlier resistance
- Ledoit-Wolf shrinkage covariance
- Optional volatility targeting
- Rebalancing frequency optimization

5.7 NCO: Nested Clustered Optimization

NCO represents the most sophisticated evolution, combining clustering with convex optimization within and across clusters:

NCO Architecture:

Two-Level Optimization:

1. **Intra-Cluster Optimization:**
   - Within each cluster, solve a convex optimization problem
   - Options: min_variance, equal_weight, equal_vol, risk_parity
   - Can apply position limits and constraints
2. **Inter-Cluster Optimization:**
   - Allocate across clusters as meta-assets
   - Options: min_variance, risk_parity, equal_weight
   - Treats each cluster as a single asset with aggregated properties

```
NCO Objective:

min w'Σw subject to:
Σw_i = 1 (within cluster)
w_min ≤ w_i ≤ w_max
Additional constraints...
```

Advantages over HRP/HERC/MHRP:

- Flexibility in optimization objectives
- Explicit constraint handling
- Better for specific risk/return targets
- Can incorporate views and predictions

5.8 Method Comparison and Selection Guide

Method Selection Matrix:

Method	Best For	Strengths	Limitations	Typical Concentration
HRP	General diversification	Stable, no parameters	Can over-allocate to low-vol assets	Moderate
HERC	Risk parity goals	Equal risk contribution	Complex implementation	Well-distributed

MHRP	Volatile markets	Balanced allocations	Less aggressive risk reduction	Balanced
NCO	Constrained optimization	Flexible, handles constraints	Parameter sensitivity	Variable

5.9 Implementation Architecture

System Workflow:

1. **Data Preparation:**
    - Fetch historical prices (configurable lookback)
    - Calculate returns matrix
    - Handle missing data and outliers
2. **Covariance Estimation:**
    - Apply Ledoit-Wolf shrinkage
    - Calculate correlation matrix (Pearson or Spearman)
    - Convert to distance matrix
3. **Hierarchical Clustering:**
    - Apply linkage method (Ward, single, complete)
    - Generate dendrogram
    - Perform quasi-diagonalization
4. **Weight Optimization:**
    - Apply selected method (HRP/HERC/MHRP/NCO)
    - Enforce constraints if applicable
    - Normalize weights to sum to 1
5. **Risk Analytics:**
    - Calculate expected return and volatility
    - Compute Sharpe ratio
    - Measure diversification metrics (HHI, effective N)
    - Generate risk contribution analysis

5.10 Risk Metrics and Performance Analysis

Portfolio Metrics

- **Expected Return:** $\mu\_p = w'\mu$
- **Volatility:** $\sigma\_p = \sqrt{(w'\Sigma w)}$
- **Sharpe Ratio:** $SR = \mu\_p / \sigma\_p$
- **Max Weight:** $\max(w\_i)$

Diversification Metrics

- **HHI:** Σw_i^2 (concentration)
- **Effective N:** 1/HHI
- **Diversification Ratio:** $\Sigma w\_i\sigma\_i / \sigma\_p$
- **Risk Parity Distance:** Deviation from equal RC

Visualization Outputs

- Donut charts of allocations
- Dendrograms of clustering
- Correlation heatmaps
- Risk contribution plots

5.11 Practical Considerations

Implementation Best Practices:

Data Requirements:

- Minimum 60 days of returns for reliable estimation
- Ideal: 180-252 days (6-12 months)
- Daily returns preferred over weekly/monthly

Parameter Guidelines:

- **Correlation Method:** Spearman for robustness, Pearson for linearity
- **Linkage Method:** Ward for balanced clusters, Single for chaining
- **Max Clusters (HERC/NCO):** $\sqrt{N}$ where N = number of assets
- **Position Limits:** Min 2%, Max 20% for institutional portfolios

Rebalancing Considerations:

- Monthly rebalancing typical for stable portfolios
- Trigger-based for volatile markets (5-10% drift threshold)
- Consider transaction costs in rebalancing frequency

5.12 Academic References and Attribution

Original Research Papers:

- **HRP:** López de Prado, M. (2016). "Building Diversified Portfolios that Outperform Out of Sample." Journal of Portfolio Management.
- **HERC:** Raffinot, T. (2017). "Hierarchical Clustering-Based Asset Allocation." Journal of Portfolio Management.
- **NCO:** López de Prado, M. (2020). "Machine Learning for Asset Managers." Cambridge University Press.
- **Shrinkage:** Ledoit, O. & Wolf, M. (2004). "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices." JMVA.

Key Innovation: These hierarchical methods represent a paradigm shift from traditional optimization by replacing unstable matrix inversion with robust machine learning techniques, providing superior out-of-sample performance and intuitive portfolio structures that align with economic relationships between assets.

Implementation Notes:

- All methods are parameter-free in their basic form (except NCO)
- Computational complexity is $O(N^2 \log N)$ vs $O(N^3)$ for traditional methods
- No need for return predictions - only covariance estimation required
- Methods are deterministic - same inputs always produce same outputs
- Can be combined with factor models and other enhancements

Complete System Documentation v3.0

This documentation covers all components of MSAS including Market Regime Detection, Multifactor Stock Scoring, Score Trend Analysis, Portfolio Selection, and Portfolio Optimization.