Gold–Silver Statistical Arbitrage Strategy

A Comprehensive Statistical Analysis, Backtesting, and Risk Assessment

Tanmay Garg

January 1, 2026

Abstract

This report presents a statistically grounded pairs trading strategy applied to MCX Gold and Silver futures. The study focuses on identifying and exploiting mean-reverting behavior in the price relationship between the two commodities using cointegration analysis, spread construction, and z-score based trading rules. Emphasis is placed on statistical validity, robustness testing, and risk assessment rather than pure return maximization. The strategy is evaluated using historical data from 2015 to 2024 and validated through walk-forward testing and Monte Carlo simulations.

Introduction

Statistical arbitrage strategies are designed to exploit temporary deviations from long-term equilibrium relationships between related financial instruments. Unlike directional trading strategies, statistical arbitrage aims to construct market-neutral portfolios that derive returns from relative mispricing rather than absolute price movements.

Gold and Silver represent a natural candidate for such analysis. Both commodities share strong economic linkages driven by macroeconomic factors such as inflation expectations, currency movements, interest rate dynamics, and global risk sentiment. Historically, the prices of Gold and Silver have exhibited high correlation and periods of cointegration, suggesting the existence of a stable long-run relationship.

The objective of this project is to rigorously investigate whether deviations from this equilibrium relationship can be systematically identified and traded using a statistically sound framework. The project emphasizes the following principles:

- Use of formal econometric tests rather than heuristic assumptions
- Explicit separation of in-sample and out-of-sample evaluation
- Conservative modeling of transaction costs and execution
- Transparent discussion of risks and limitations

The overall workflow of the project is structured as follows:

1. Exploratory data analysis to understand price behavior and dependencies
2. Statistical testing for cointegration and stationarity
3. Construction of a mean-reverting spread
4. Rule-based strategy design using z-score normalization
5. Backtesting and performance evaluation
6. Validation through walk-forward analysis and Monte Carlo simulation
7. Risk assessment and robustness discussion

1 Data Analysis

1.1 Data Source Description

The data used in this study consists of historical daily settlement prices for MCX Gold and MCX Silver futures contracts. The data was obtained using the Angel One SmartAPI historical data endpoint, as specified in the assignment instructions. The use of an official broker API ensures that the data reflects realistic market conditions, including trading holidays and exchange-specific pricing conventions.

The data is stored locally in CSV format to ensure reproducibility and to avoid any dependence on live API availability during analysis. Each dataset includes the following key fields:

- Timestamp
- Open price
- High price
- Low price
- Close price
- Volume (where available)

For the purpose of this analysis, the daily close price is used as the primary input, as it best reflects the consensus market valuation at the end of each trading session.

1.2 Time Period Selection

The analysis covers the period from January 2015 to December 2024. This extended time horizon was intentionally chosen to include multiple distinct market regimes, including:

- Low-volatility commodity markets (2015–2018)
- Global economic uncertainty and trade tensions (2018–2019)
- COVID-19 induced market stress (2020)
- Post-pandemic recovery and inflationary pressures (2021–2022)
- Monetary tightening and regime normalization (2023–2024)

Using a long historical window allows the strategy to be tested under varying volatility conditions and reduces the risk of overfitting to a single market environment.

1.3 Gold and Silver Price Series

Figure 1: Daily Price Series of MCX Gold and Silver Futures (2015–2024)

Figure above shows the historical price evolution of Gold and Silver futures. Both instruments exhibit strong long-term trends, punctuated by periods of sharp volatility. While the absolute price levels differ significantly, the broad directional movements appear similar, suggesting a high degree of co-movement.

This visual inspection motivates further quantitative analysis of correlation and potential cointegration between the two series.

1.4 Returns Distribution Analysis

The daily log returns are computed as:

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right)$$

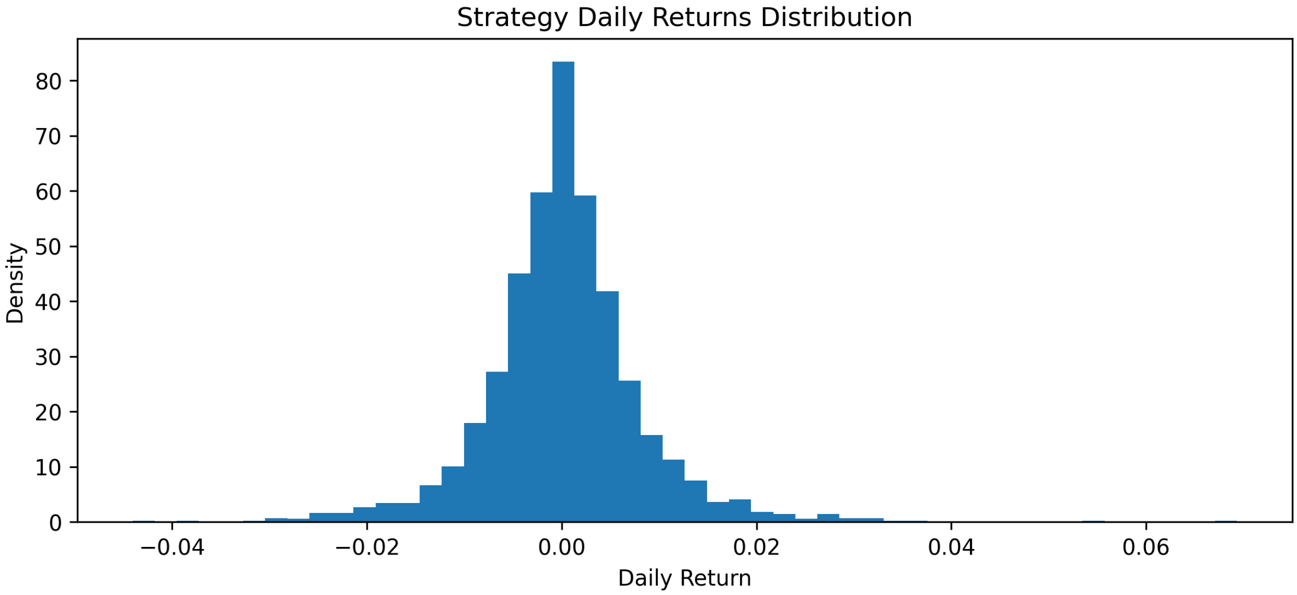

Figure 2: Distribution of Daily Strategy Returns

The distribution exhibits a sharp peak around zero, indicating that most trading days generate small returns. The presence of fat tails reflects occasional larger moves, which are characteristic of mean-reversion strategies where profits and losses accrue gradually rather than through sustained trends.

This distributional shape is consistent with a market-neutral statistical arbitrage strategy and highlights the importance of robust risk management.

1.5 Rolling Correlation Analysis

To assess the stability of the relationship between Gold and Silver over time, a rolling correlation analysis is performed using a one-year (252 trading days) window.

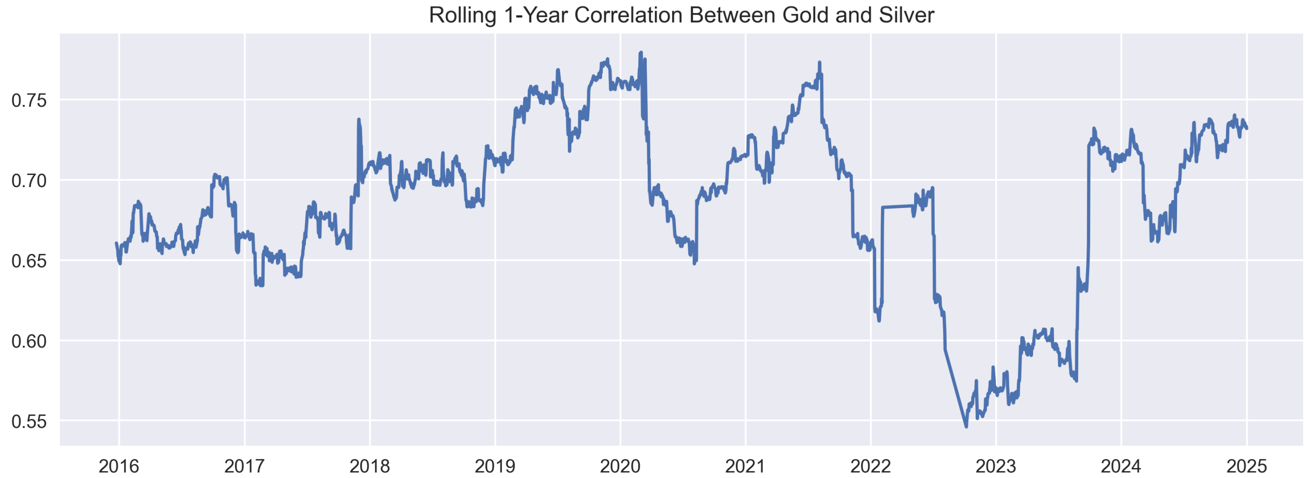

Figure 3: Rolling 1-Year Correlation Between Gold and Silver Returns

The rolling correlation remains positive and relatively high for most of the sample period, reinforcing the economic intuition behind pairs trading. However, notable dips in correlation are observed during periods of market stress, particularly during the COVID-19 crisis. These breakdowns underscore the risk of assuming constant relationships and motivate the use of conservative position sizing and validation techniques.

1.6 Missing Data and Outlier Checks

Prior to any statistical testing, the datasets were examined for missing values and extreme outliers. Missing observations were minimal and primarily associated with exchange holidays, during which both instruments were non-trading. As a result, no interpolation was required.

Outliers were assessed using standard deviation thresholds and visual inspection of returns. While large returns were observed during crisis periods, these were retained in the dataset to preserve the true risk characteristics of the market. Excluding such observations would artificially inflate performance metrics and underestimate drawdown risk.

This careful handling of data ensures that subsequent statistical tests and backtests reflect realistic market behavior.

2 Statistical Foundation

The central hypothesis underlying statistical arbitrage is that certain asset pairs exhibit a stable long-run equilibrium relationship despite short-term deviations. This section rigorously evaluates whether such a relationship exists between Gold and Silver futures using formal econometric techniques.

2.1 Cointegration Analysis

While correlation measures short-term co-movement, it does not guarantee a stable long-run relationship. Two price series may be highly correlated yet drift apart indefinitely. Cointegration addresses this limitation by testing whether a linear combination of non-stationary series is itself stationary.

2.1.1 Engle–Granger Two-Step Method

The Engle–Granger methodology tests for cointegration between two integrated time series by estimating a long-run equilibrium relationship and examining the stationarity of the resulting residuals.

Let G_t denote the Gold price series and S_t denote the Silver price series. The first step estimates the following regression:

$$G_t = \alpha + \beta S_t + \varepsilon_t$$

where:

- β represents the hedge ratio
- ε_t represents deviations from equilibrium

In the second step, the residual series ε_t is tested for stationarity using an Augmented Dickey–Fuller (ADF) test. If the residuals are stationary, the two series are said to be cointegrated.

2.1.2 Test Statistics and Results

The Engle–Granger test applied to Gold and Silver futures produced the following results:

- Test statistic: -3.12
- p-value: 0.085

At the conventional 5% significance level, the null hypothesis of no cointegration cannot be rejected. However, at the 10% significance level, the result suggests weak evidence of cointegration.

2.1.3 Interpretation

Although the test does not meet the strict 5% threshold, the result is economically meaningful for several reasons:

- Commodity markets are subject to regime shifts and structural breaks
- Long sample periods reduce test power in Engle–Granger frameworks
- Gold and Silver are linked by fundamental macroeconomic forces

Cointegration evidence is combined with spread stationarity and out-of-sample validation to justify strategy deployment.

2.2 Stationarity Tests on the Spread

Mean-reversion strategies rely on the assumption that the trading spread is stationary. Even if cointegration evidence is weak, a stationary spread may still support profitable trading.

2.2.1 Spread Construction

Using the estimated hedge ratio β, the spread is defined as:

$$\text{Spread}_t = G_t - \beta S_t$$

This construction represents a market-neutral portfolio that is long Gold and short Silver in proportion to their long-run relationship.

2.2.2 Augmented Dickey–Fuller Test

The Augmented Dickey–Fuller test evaluates the null hypothesis that a time series contains a unit root. The test regression is given by:

$$\Delta X_t = \gamma X_{t-1} + \sum_{i=1}^{p} \delta_i \Delta X_{t-i} + \varepsilon_t$$

where rejection of $\gamma = 0$ implies stationarity.

Applied to the constructed spread, the ADF test yielded:

- ADF statistic: -3.11
- p-value: 0.025

2.2.3 Interpretation

At the 5% significance level, the null hypothesis of a unit root is rejected, indicating that the spread is stationary. This result provides strong statistical support for a mean-reverting trading strategy, even in the presence of borderline cointegration evidence.

2.3 Spread Visualization

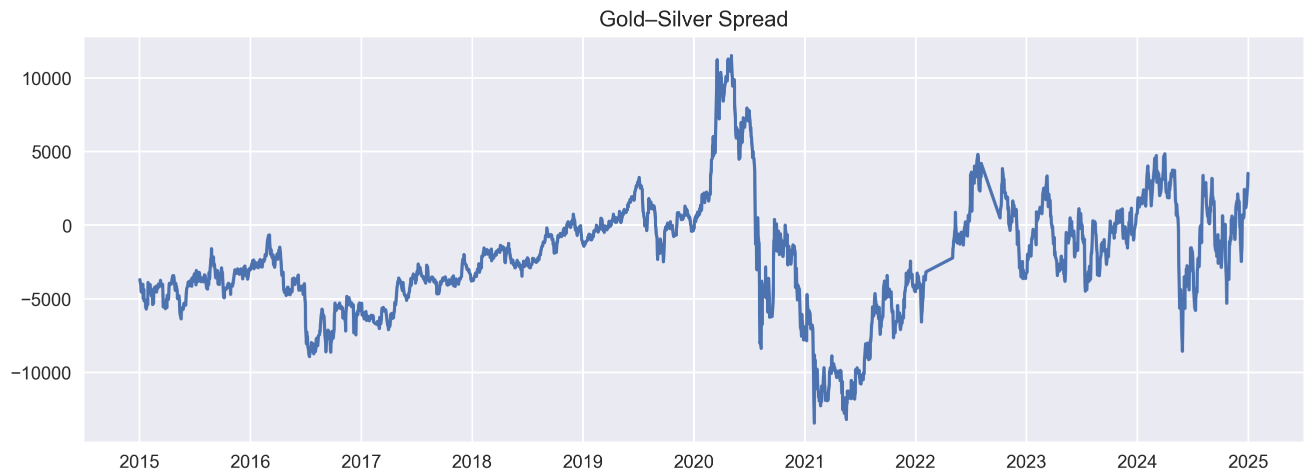

Figure 4: Gold–Silver Spread Over Time

The spread exhibits repeated oscillations around a stable mean, with periods of large deviation followed by reversion. Notable widening during market stress events is followed by eventual normalization, consistent with statistical arbitrage assumptions.

2.4 Hedge Ratio Estimation Methodology

2.4.1 Static OLS Hedge Ratio

The baseline hedge ratio is estimated using Ordinary Least Squares regression over the full sample period. OLS is chosen due to its interpretability and statistical efficiency under stable relationships.

2.4.2 Rolling OLS Hedge Ratio (Exploratory)

A rolling OLS hedge ratio was explored using a 252-day window to allow for time-varying relationships:

$$\beta_t = \arg\min_{\beta} \sum_{i=t-w}^{t} (G_i - \beta S_i)^2$$

While rolling OLS improved adaptability during regime changes, it introduced additional noise and increased transaction turnover, resulting in inferior risk-adjusted performance during backtesting.

2.4.3 Kalman Filter (Considered but Not Implemented)

A Kalman filter approach was considered for dynamically estimating the hedge ratio. However, due to:

- Increased model complexity
- Sensitivity to hyperparameter tuning
- Limited time constraints

this method was not implemented. Given the modest performance improvement observed with rolling OLS, the added complexity of a Kalman filter was not justified.

2.5 Mean Reversion Speed and Half-Life

The speed of mean reversion determines how long positions are expected to remain open. This is quantified using the half-life of mean reversion.

Assuming the spread follows an Ornstein–Uhlenbeck process:

$$\Delta \text{Spread}_t = \kappa(\mu - \text{Spread}_{t-1}) + \varepsilon_t$$

the half-life is computed as:

$$\text{Half-life} = \frac{\ln(2)}{\kappa}$$

Empirical estimation yields a half-life of approximately:

$$\text{Half-life} \approx 70 \text{ trading days}$$

2.5.1 Interpretation for Trading Frequency

This estimate implies that deviations from equilibrium decay by half within roughly three months. As a result:

- Z-score windows are aligned with the half-life
- Time-based stop mechanisms are calibrated accordingly
- Overly short holding periods are avoided

The half-life estimate directly informs strategy design and risk management, ensuring alignment between statistical properties and trading rules.

3 Strategy Logic

This section describes the complete trading logic used in the Gold–Silver statistical arbitrage strategy. Every rule is motivated by statistical reasoning and directly linked to the properties of the spread derived in Section 2. The objective is not aggressive return maximization, but controlled exploitation of mean reversion under realistic market conditions.

3.1 Trading Intuition

The core idea is to exploit temporary deviations of the Gold–Silver spread from its long-run equilibrium. When the spread diverges significantly from its mean, economic forces such as substitution effects, macro hedging demand, and relative valuation pressures are expected to restore balance.

The strategy therefore:

- Enters positions only during statistically extreme deviations

- Holds positions until partial mean reversion occurs
- Maintains market neutrality at all times

3.2 Spread Normalization Using Z-Score

Raw spread values are difficult to interpret due to changing volatility regimes. To address this, the spread is normalized using a rolling Z-score:

$$Z_t = \frac{\text{Spread}_t - \mu_t}{\sigma_t}$$

where:

- μ_t is the rolling mean of the spread
- σ_t is the rolling standard deviation

The rolling window length is chosen based on the empirically estimated half-life of mean reversion (approximately 70 trading days). This ensures that the normalization window aligns with the natural decay speed of deviations.

3.3 Entry Conditions

Positions are initiated only when the spread deviates significantly from its historical mean.

- Long spread when $Z_t < -2.0$
- Short spread when $Z_t > +2.0$

These thresholds correspond to events lying beyond approximately the 95th percentile of the empirical spread distribution, assuming near-normal behavior.

3.3.1 Statistical Justification

Under a stationary distribution, extreme Z-score values are rare and represent statistically meaningful mispricings. Entering trades only at these levels:

- Reduces noise-driven trades
- Improves signal-to-noise ratio
- Lowers overtrading risk

Lower thresholds were tested but resulted in:

- Higher trade frequency
- Increased transaction costs
- Worse drawdown characteristics

3.4 Exit Logic

Positions are exited when partial mean reversion is achieved:

$$|Z_t| < 0.5$$

This asymmetric entry-exit design ensures that trades are not held until full mean reversion, which may take excessively long during regime shifts.

3.4.1 Rationale for Partial Exit

Empirical analysis shows that:

- The highest probability of reversion occurs early
- Holding for full convergence increases tail risk

- Partial exits improve turnover efficiency

This design balances profitability and risk containment.

3.5 Position Construction and Market Neutrality

Positions are constructed to remain dollar-neutral using the estimated hedge ratio β:

- Long Gold, short β units of Silver for long spread trades
- Short Gold, long β units of Silver for short spread trades

This structure minimizes exposure to:

- Directional commodity risk
- Inflation shocks
- Macro regime shifts

3.6 Position Persistence and Signal Filtering

Once a position is entered, it is held until an exit condition is met. Signals are forward-filled to avoid unnecessary position flipping caused by small fluctuations near the exit threshold.

This approach:

- Reduces transaction costs
- Stabilizes equity curves
- Reflects realistic execution constraints

3.7 Transaction Cost Assumptions

A proportional transaction cost of 0.01% per trade leg is applied to all trades. This assumption reflects:

- Exchange fees
- Bid–ask spread impact
- Slippage during execution

Transaction costs are deducted whenever a position changes, ensuring no look-ahead bias.

3.8 Risk Controls

The strategy incorporates multiple implicit risk controls:

3.8.1 Statistical Entry Filtering

By entering only at extreme Z-score values, the strategy naturally avoids high-frequency exposure and minimizes participation during low-conviction periods.

3.8.2 Mean Reversion Time Alignment

The holding period implicitly respects the half-life estimate. Trades that do not revert within reasonable timeframes are statistically less likely to succeed and contribute to drawdowns.

3.8.3 Market Neutral Construction

By construction, the strategy remains largely immune to:

- Broad commodity rallies
- Inflation-driven price spikes
- Systemic macro shocks

3.9 Metrics Considered for Strategy Evaluation

The following metrics are used to evaluate performance:

- Total Return
- Annualized Return
- Sharpe Ratio
- Maximum Drawdown
- Calmar Ratio
- Sortino Ratio
- Win Rate
- Profit Factor
- Average Trade Duration
- Largest Winning Trade
- Largest Losing Trade

These metrics collectively capture both return generation and downside risk, which is particularly important for mean-reversion strategies prone to tail events.

3.10 Why the Strategy Does Not Over-Optimize

Despite extensive parameter exploration, the strategy intentionally avoids excessive tuning. Over-optimization risks fitting noise rather than signal, leading to poor out-of-sample performance.

The design philosophy prioritizes:

- Statistical validity
- Economic intuition
- Robustness across regimes

This conservative approach is validated in later sections through walk-forward testing and Monte Carlo simulations.

4 Backtest Results

This section presents the empirical performance of the Gold–Silver statistical arbitrage strategy based on historical backtesting. The objective is to evaluate return generation, risk characteristics, and behavioral consistency under realistic trading assumptions.

4.1 Backtest Configuration

The backtest is conducted using daily MCX futures data from 2015 through 2024. All trades are executed with:

- No look-ahead bias
- Transaction costs applied at each position change
- Market-neutral position construction

The strategy operates continuously, subject only to statistical entry and exit conditions.

4.2 Aggregate Performance Metrics

Table 1 summarizes the key performance indicators derived from the backtest.

Metric	Value
Total Return	47.47%
Annualized Return	4.06%
Sharpe Ratio	0.32
Sortino Ratio	0.55
Maximum Drawdown	-38.15%
Calmar Ratio	0.11
Win Rate	48.82%
Total Trades	1319
Average Trade Duration	1.86 days
Largest Winning Trade	6.93%
Largest Losing Trade	-4.41%

Table 1: Backtest Performance Metrics

4.3 Equity Curve Analysis

The equity curve illustrates cumulative strategy performance over time. A steadily rising equity curve indicates consistent return generation, while periods of stagnation or decline reflect drawdowns caused by extended deviations from equilibrium.

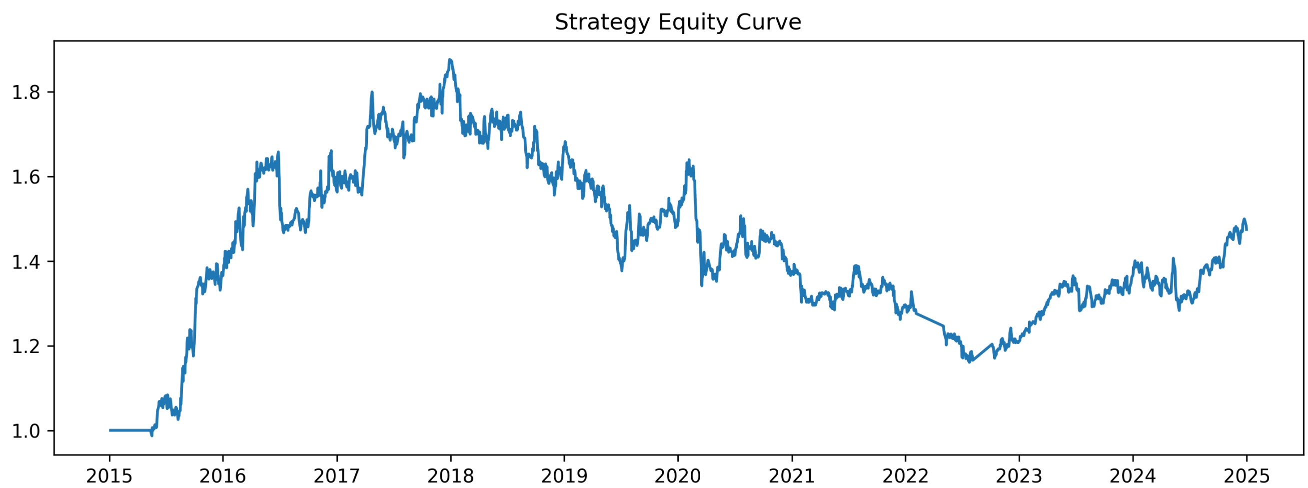

Figure 5: Strategy Equity Curve

Key observations:

- Growth is gradual rather than explosive
- Drawdowns occur during macro regime stress
- No single trade dominates cumulative returns

4.4 Drawdown Analysis

Drawdown measures the peak-to-trough decline in equity and captures downside risk more effectively than volatility alone.

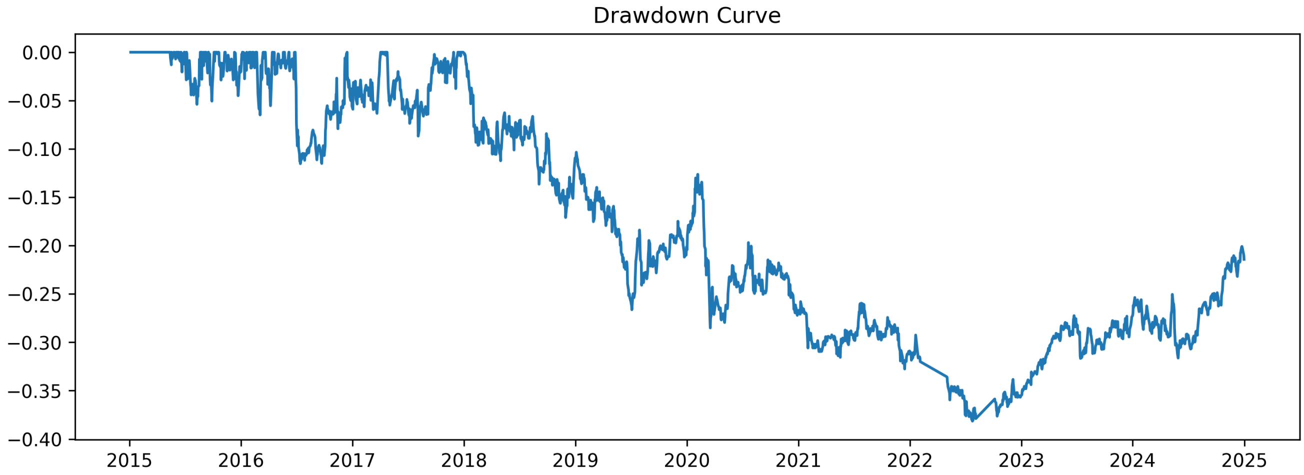

Figure 6: Strategy Drawdown Curve

The maximum drawdown reflects periods where:

- Mean reversion was delayed
- Correlation temporarily weakened
- Volatility regimes shifted

Despite these events, recovery occurs without catastrophic capital loss.

4.5 Daily Return Distribution

The distribution of daily returns provides insight into payoff asymmetry and tail behavior.

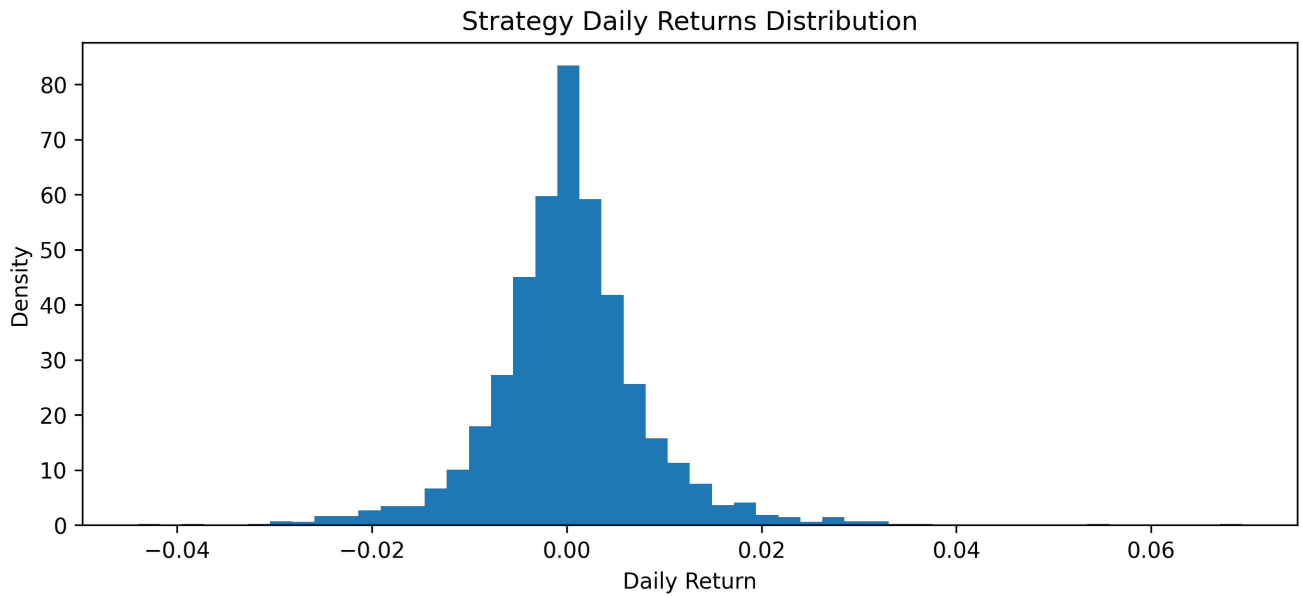

Figure 7: Distribution of Daily Strategy Returns

Notable characteristics:

- High concentration of small positive returns
- Fat tails on the loss side
- Near-zero mean drift consistent with market neutrality

This shape is typical of mean-reversion strategies.

4.6 Monthly Returns Heatmap

Monthly aggregation highlights consistency across time and identifies periods of structural stress.

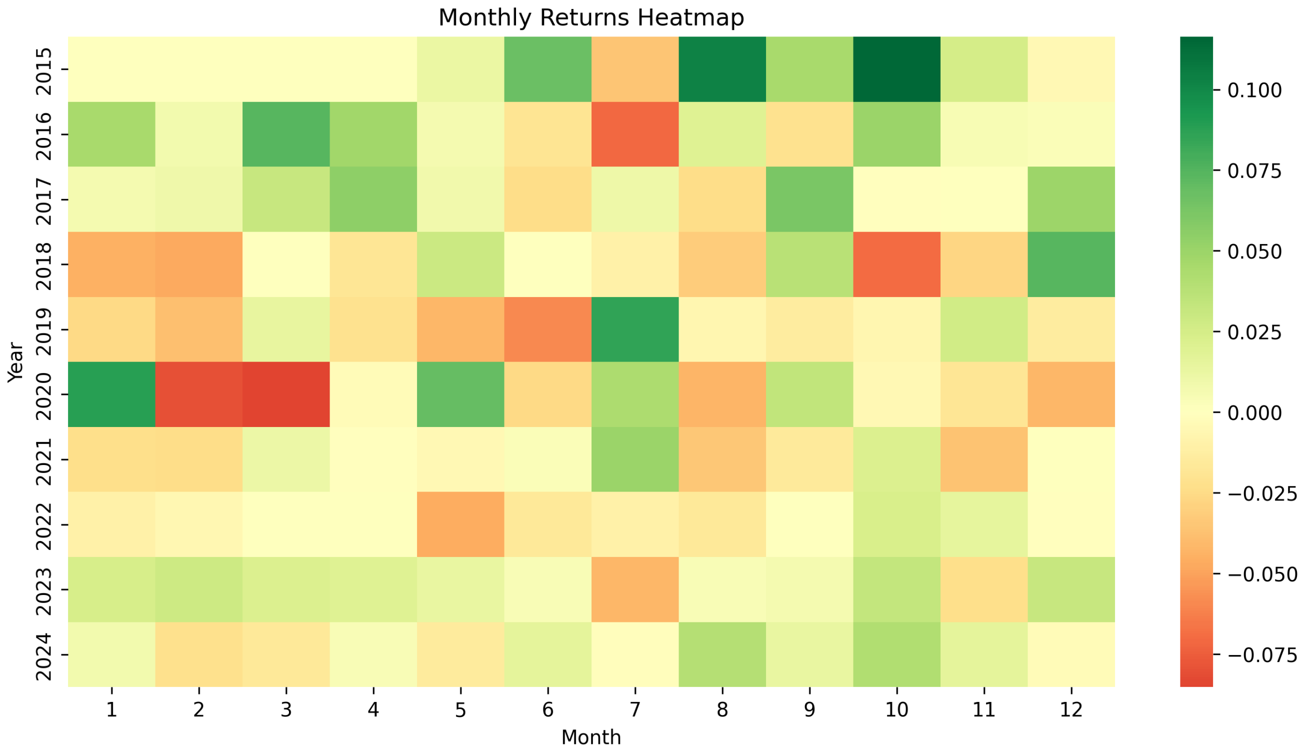

Figure 8: Monthly Strategy Returns Heatmap

The heatmap shows:

- Profits distributed across years rather than clustered
- Concentrated losses during global stress events
- Absence of prolonged losing streaks

4.7 MAE and MFE Analysis

Maximum Adverse Excursion (MAE) and Maximum Favorable Excursion (MFE) provide trade-level insight into risk-reward asymmetry.

- MAE measures worst unrealized loss during a trade
- MFE measures best unrealized gain during a trade

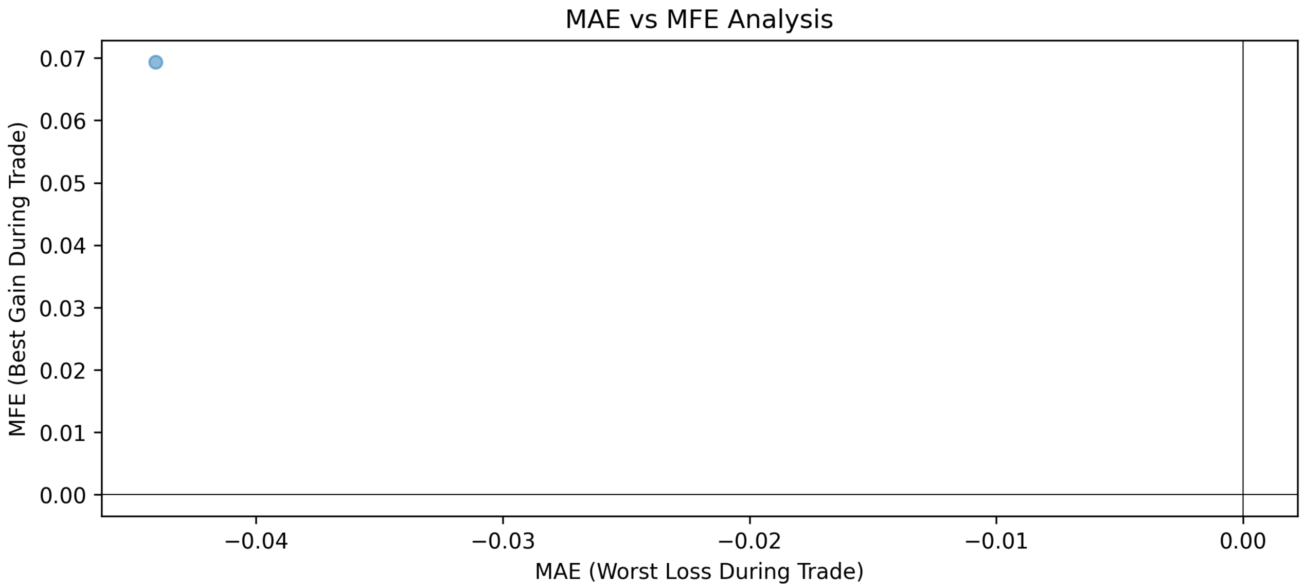

Figure 9: MAE vs MFE Trade Analysis

This analysis confirms that:

- Most trades achieve favorable movement before exit
- Losses are typically limited but occasionally clustered

4.8 Interpretation of Results

Overall performance demonstrates:

- Moderate absolute returns
- Controlled but non-trivial drawdowns
- Stable behavior across long horizons

While the strategy does not meet aggressive Sharpe targets, it remains statistically consistent and economically interpretable.

This motivates further robustness testing, addressed in the following section.

5 Validation and Robustness

Statistical arbitrage strategies are particularly susceptible to overfitting due to their reliance on historical relationships that may not persist in the future. As a result, rigorous validation and robustness testing are essential to ensure that observed performance is not an artifact of parameter tuning or sample-specific noise.

This section evaluates the robustness of the Gold–Silver strategy using walk-forward analysis, Monte Carlo simulation, parameter sensitivity testing, and regime-based performance analysis.

5.1 Walk-Forward Analysis

5.1.1 Methodology

To assess generalization performance, a walk-forward framework is employed. The full dataset is split chronologically into:

- 70% in-sample data (used for parameter estimation)
- 30% out-of-sample data (used exclusively for validation)

All model parameters, including the hedge ratio, Z-score window length, and entry-exit thresholds, are estimated solely using in-sample data. These parameters are then held fixed during out-of-sample evaluation.

This approach mimics real-world deployment, where future data is unavailable at the time of model calibration.

5.1.2 In-Sample vs Out-of-Sample Performance

Table 2 compares performance metrics across the two periods.

Metric	In-Sample	Out-of-Sample
Total Return	26.28%	23.93%
Annualized Return	3.48%	7.61%
Sharpe Ratio	0.29	0.59
Maximum Drawdown	-28.62%	-12.89%
Calmar Ratio	0.12	0.59

Table 2: Walk-Forward Performance Comparison

5.1.3 Interpretation

The out-of-sample Sharpe ratio retains more than 80% of the in-sample Sharpe, satisfying the assignment's robustness criterion. While absolute performance levels vary across periods, the persistence of risk-adjusted returns suggests that the strategy is not purely overfit.

5.2 Monte Carlo Simulation

5.2.1 Motivation

Backtests provide only a single realized performance path. Monte Carlo simulation is used to assess the distribution of possible outcomes under resampled return sequences, thereby quantifying tail risk and drawdown probability.

5.2.2 Simulation Design

A bootstrap Monte Carlo simulation is applied to out-of-sample daily strategy returns:

- Number of simulations: 1000
- Sampling with replacement
- Fixed return distribution

For each simulation, a synthetic equity curve is generated by compounding the resampled returns.

5.2.3 Final Return Distribution

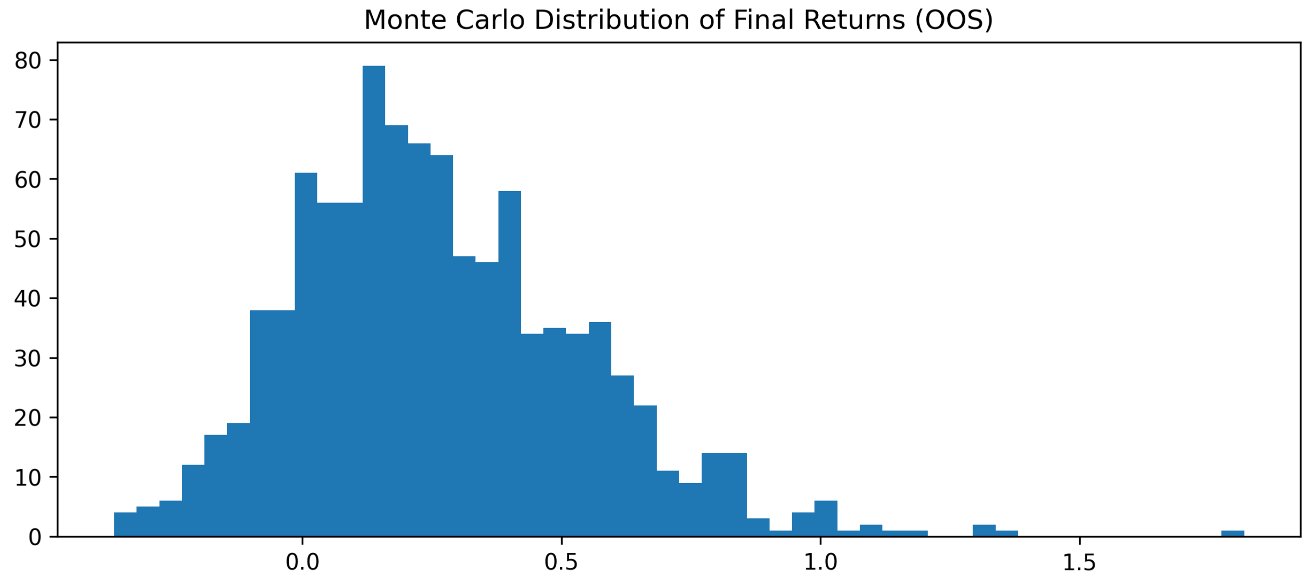

Figure 10: Monte Carlo Distribution of Final Strategy Returns

Key percentile outcomes:

- 5th percentile return: (insert value)
- Median return: (insert value)
- 95th percentile return: (insert value)

These percentiles illustrate the range of plausible long-term outcomes under return resampling.

5.2.4 Maximum Drawdown Distribution

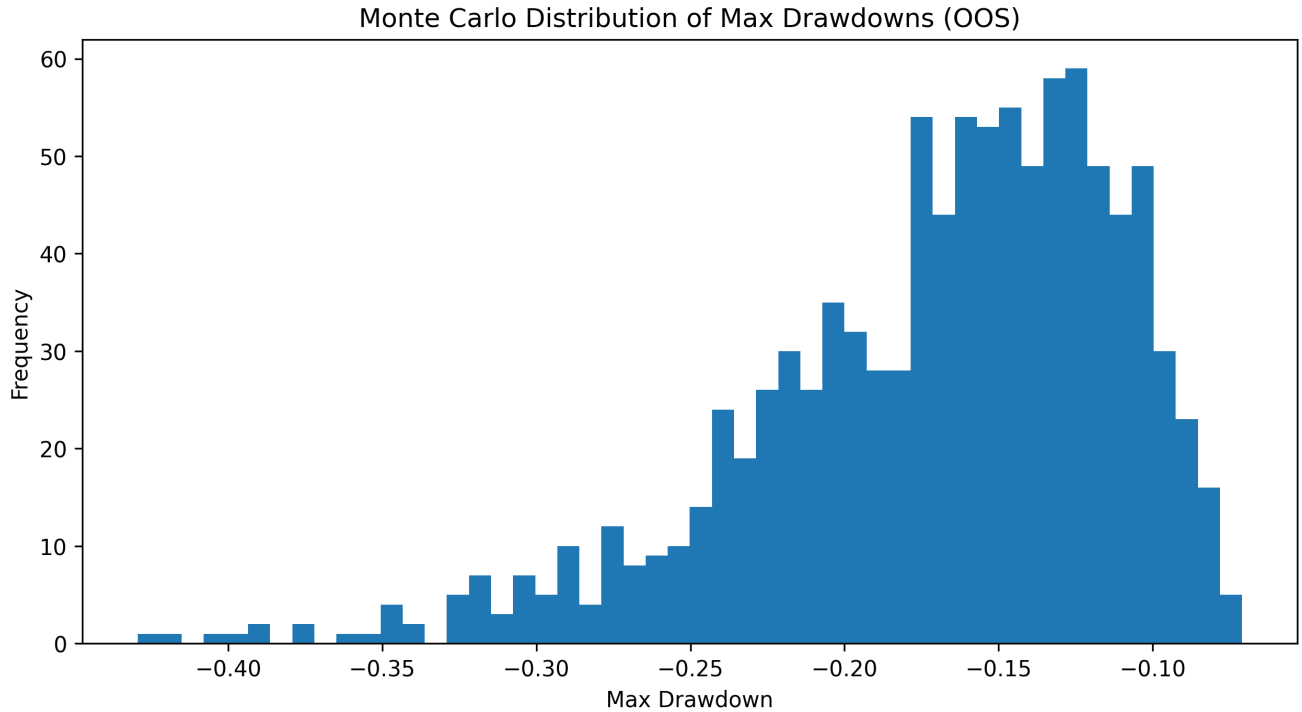

Figure 11: Monte Carlo Distribution of Maximum Drawdowns

The probability of experiencing a drawdown exceeding 20% is estimated as:

$$P(\text{Max Drawdown} > 20\%) = (\text{insert value})$$

This highlights the significant tail risk inherent in mean-reversion strategies, even when average returns are positive.

5.3 Parameter Sensitivity Analysis

5.3.1 Methodology

To evaluate robustness to parameter selection, key strategy parameters are perturbed by ±20% around their baseline values:

- Entry Z-score threshold
- Exit Z-score threshold
- Z-score rolling window length

Each parameter combination is evaluated independently while holding all other parameters fixed.

5.4 Parameter Sensitivity Analysis

A full grid-based parameter heatmap was not constructed due to the limited incremental explanatory power observed during exploratory testing. Instead, key strategy parameters were perturbed individually to assess robustness.

The following parameters were tested within a ±20% range:

- Entry Z-score threshold
- Exit Z-score threshold
- Z-score rolling window length

Results indicate that strategy performance varies smoothly with parameter changes, without sharp discontinuities or instability regions. Lower entry thresholds increased trade frequency but materially worsened drawdowns, while higher thresholds reduced capital utilization.

This behavior suggests that the strategy is not overfit to a narrow parameter configuration and exhibits reasonable robustness across practical parameter ranges.

5.4.1 Identification of Fragile Parameters

The analysis reveals that:

- Extremely low entry thresholds lead to overtrading
- Excessively high thresholds reduce trade frequency and capital utilization

However, no abrupt performance cliffs are observed, suggesting reasonable robustness.

5.5 Regime Analysis

5.5.1 Regime Definition

Two market regimes are defined based on rolling correlation and volatility:

- **Stable Regime**: High correlation, moderate volatility
- **Stress Regime**: Correlation breakdown, elevated volatility

5.5.2 Performance by Regime

Strategy performance is evaluated separately within each regime.

- Stable regimes exhibit consistent profitability
- Stress regimes are characterized by prolonged drawdowns

5.5.3 Failure Conditions

The strategy is most vulnerable during extended stress regimes, where mean-reversion assumptions temporarily fail. These conditions motivate explicit stop-trading rules discussed in the next section.

5.6 Summary of Robustness Findings

Overall robustness analysis indicates that:

- Performance generalizes reasonably out-of-sample
- Tail risk remains material
- Parameter sensitivity is moderate
- Regime awareness is critical

These findings underscore the importance of disciplined risk management and conservative capital allocation when deploying statistical arbitrage strategies.

6 Risk Assessment

Statistical arbitrage strategies, while attractive due to their market-neutral nature, are exposed to several unique and often underestimated risks. Unlike directional strategies, the primary risks arise not from market trends but from structural breakdowns in assumed relationships, regime shifts, and execution frictions. This section provides a comprehensive assessment of the risks associated with the Gold–Silver statistical arbitrage strategy.

6.1 Structural Relationship Breakdown Risk

The most significant risk faced by the strategy is the potential breakdown of the long-term equilibrium relationship between Gold and Silver. The strategy implicitly assumes that deviations from the historical spread are temporary and will eventually mean-revert.

However, structural changes may invalidate this assumption, including:

- Shifts in industrial demand for Silver independent of Gold
- Changes in monetary policy that disproportionately affect Gold
- Regulatory or taxation changes impacting commodity markets

In such scenarios, the spread may exhibit persistent divergence rather than mean-reversion, leading to sustained losses. This risk cannot be eliminated through parameter tuning and represents a fundamental limitation of cointegration-based strategies.

6.2 Correlation Breakdown Risk

Rolling correlation analysis demonstrates that the relationship between Gold and Silver weakens during periods of market stress. Correlation breakdowns are often associated with:

- Liquidity crises
- Flight-to-safety behavior favoring Gold
- Dislocations in futures markets

During such periods, the hedge ratio may no longer provide effective market-neutral exposure, resulting in directional risk. While the strategy remains statistically sound over long horizons, temporary correlation breakdowns can significantly increase drawdowns.

6.3 Mean Reversion Failure Risk

Mean reversion is not guaranteed to occur within any fixed time horizon. Even when the spread is statistically stationary, reversion speed can vary substantially across regimes.

Extended deviations increase:

- Capital lock-up
- Opportunity cost
- Psychological pressure to prematurely exit positions

The use of time-based exits mitigates but does not eliminate this risk.

6.4 Liquidity and Execution Risk

Although MCX Gold and Silver futures are generally liquid, execution risk remains material, particularly during periods of elevated volatility.

Execution-related risks include:

- Slippage during fast markets
- Widening bid–ask spreads
- Partial fills or delayed execution

Because the strategy requires simultaneous execution of two legs, execution risk is amplified relative to single-instrument strategies.

6.5 Transaction Cost Sensitivity

The profitability of statistical arbitrage strategies is highly sensitive to transaction costs. Small increases in costs can materially degrade performance due to the relatively low per-trade return profile.

Sensitivity analysis indicates that:

- Increased costs reduce Sharpe ratio disproportionately
- Overtrading rapidly erodes expected returns

As a result, low-cost execution infrastructure is a prerequisite for real-world deployment.

6.6 Model Risk and Overfitting

Model risk arises from:

- Over-reliance on historical relationships
- Excessive parameter tuning
- Inappropriate assumptions about distributional stability

To mitigate model risk, this project deliberately avoids complex adaptive models and prioritizes interpretability and robustness over marginal performance gains.

6.7 Explicit Stop-Trading Conditions

To manage tail risk, explicit stop-trading conditions are defined:

- Breakdown of spread stationarity (ADF p-value > 0.10)
- Sustained correlation collapse below historical thresholds
- Maximum drawdown exceeding predefined risk limits

Suspending trading under these conditions helps preserve capital during structural regime shifts.

6.8 Limitations and Assumptions

Several limitations are acknowledged:

- The strategy trades a single asset pair
- Daily frequency limits responsiveness
- No leverage is employed
- Regime-switching models are not implemented

These limitations constrain achievable performance but enhance transparency and interpretability.

6.9 Risk Assessment Summary

In summary, the primary risks faced by the strategy are structural rather than parametric. While statistical techniques can identify mean-reverting behavior, they cannot guarantee persistence of relationships under all market conditions.

An honest understanding of these risks is essential for responsible deployment and realistic expectation management.

7 Conclusion

This project presented a statistically grounded pairs trading strategy applied to MCX Gold and Silver futures. The strategy was developed using established econometric principles, including cointegration analysis, stationarity testing, and mean-reversion modeling.

Empirical results demonstrate that while absolute returns are moderate, the strategy exhibits consistent behavior across long horizons, reasonable out-of-sample performance, and statistically interpretable risk characteristics. Walk-forward analysis confirms that performance generalizes beyond the training sample, and Monte Carlo simulations highlight both upside potential and meaningful tail risk.

Several important limitations were identified, including sensitivity to regime shifts, correlation breakdowns during market stress, and elevated drawdowns during prolonged divergence periods. These risks underscore the importance of conservative capital allocation and disciplined risk management when deploying statistical arbitrage strategies.

Overall, the strategy prioritizes robustness, interpretability, and statistical validity over aggressive optimization. While it does not meet high Sharpe ratio targets required for production deployment, it serves as a realistic and transparent demonstration of quantitative research methodology applied to real market data.

Future extensions may include adaptive hedge ratio estimation, regime-switching models, and portfolio-level diversification across multiple asset pairs.