Business Report

SILVANA ORTIZ

MICROESTRUCTURA
Y SISTEMAS DE TRADING

INDEX

OBJECTIVE

The primary objective of this project is to design, implement, and evaluate a statistical arbitrage strategy based on pairs trading principles. The project seeks to leverage econometric and quantitative methods to identify pairs of financial assets whose prices move together in the long term, exhibiting cointegration.

A central component of the strategy involves the implementation of Kalman filters to estimate dynamic hedge ratios in real time. By framing the Kalman filter as a sequential decision process following Powell’s Sequential Decision Analysis framework, in this project I will explore how state estimation, uncertainty, and adaptive decision making can be integrated into a live trading system.

The resulting model will be used to develop a market neutral trading strategy that aims to profit from short term mean reversion while maintaining neutrality to overall market movements.

STRATEGY DESCRIPTION AND RATIONALE

Overview of pairs trading approach

Pairs trading is a market neutral statistical arbitrage strategy that exploits temporary mispricings between two historically related assets. The core idea is to identify pairs of securities whose prices move together in the long run but may temporarily diverge due to short term inefficiencies.
When the spread between the two assets deviates significantly from its historical equilibrium, the strategy takes offsetting long and short positions long the undervalued asset and short the overvalued one anticipating mean reversion.
The market neutral aproach allows profits to be driven primarily by relative price movements rather than broad market trends.
Pairs trading is particularly attractive in volatile or sideways markets, where relative value opportunities are more pronounced.

The general workflow of the strategy involves:

- Identifying statistically significant cointegrated pairs among a universe of assets.
- Estimating and dynamically updating the hedge ratio that maintains neutrality.
- Constructing a spread (the residual from the cointegrating relationship) and generating trading signals based on deviations from its mean or equilibrium level.
- Implementing entry and exit thresholds that balance profitability against transaction costs and execution risk.

STRATEGY DESCRIPTION AND RATIONALE

Cointegration indicates arbitrage opportunity

Cointegration implies that although two price series may be non-stationary individually, a linear combination of them is stationary, indicating a stable long run equilibrium relationship. When two assets are cointegrated, deviations from this equilibrium are expected to be temporary, driven by transitory shocks or liquidity imbalances rather than fundamental shifts.

From an arbitrage perspective:

- If the spread widens beyond its historical mean, the strategy shorts the outperforming asset and buys the underperforming one.
- As prices revert to their equilibrium relationship, the spread narrows, and both positions can be closed at a profit.

Cointegration serves as statistical evidence of equilibrium pricing, where deviations represent potential mispricings rather than structural divergence. It makes sure that the convergence trade has a meaningful economic basis rather than relying purely on correlation, which can be unstable over time.

STRATEGY DESCRIPTION AND RATIONALE

Kalman filter use in dynamic hedging

Traditional pairs trading assumes a static hedge ratio typically estimated via ordinary least squares (OLS).
However, in real markets, relationships between assets evolve due to changes in market regimes, liquidity, and fundamental conditions. A static model may therefore misestimate the true hedge ratio, leading to suboptimal exposure and increased risk.
The Kalman filter provides a dynamic and adaptive framework for estimating time varying hedge ratios.
It models the cointegrating relationship as a state space model, where:

- The underlying hedge ratio evolves as a latent state variable following a stochastic process.
- Observed prices update the model sequentially as new data arrives.

This formulation aligns naturally with Powell's Sequential Decision Analysis framework, which treats each update as a real-time decision process involving state estimation and control actions (portfolio adjustments). The Kalman filter optimally combines prior estimates and new observations, minimizing estimation error variance in a recursive manner.
This dynamic adaptation allows the strategy to remain effective even under non stationary market conditions, ensuring the hedge ratio reflects the most current information and mitigating exposure to changing relationships between assets.

STRATEGY DESCRIPTION AND RATIONALE

Expected market conditions for strategy success

Under the following conditions, the Kalman filter based dynamic hedging mechanism can optimally adapt to gradual shifts while preserving the strategy's market neutrality and mean reversion exploitation.

- ***Mean-reverting environments:*** The approach relies on temporary deviations from equilibrium that eventually converge. Excessively trending or structurally shifting markets may break the cointegrating relationship, reducing profitability.
- ***High liquidity and low transaction costs:*** Since pairs trading involves frequent adjustments, the success of the strategy depends on tight spreads, sufficient depth, and manageable execution costs.
- ***Stable macroeconomic regimes:*** Stable conditions help preserve the long-run cointegration relationships between assets. Regime changes, such as monetary policy shocks or sectoral revaluations, can weaken the equilibrium link.
- ***Moderate volatility:*** Some volatility is necessary to create spread deviations large enough to trade on, but extreme volatility can lead to persistent disequilibria and higher risk of stop-outs.

PAIR SELECTION METHODOLOGY

Correlation screening criteria and results

The first step in identifying viable pairs for a statistical arbitrage strategy is to screen the universe of candidate assets for high historical correlation. This ensures that the selected pairs share a similar price movement pattern.

The methodology followed in this project to make sure this happens was:

- Daily price data over 15 years was collected for the chosen assets.
- Daily log returns were calculated for each asset.
- A rolling correlation matrix was computed using a 252 trading day window.
- For each pair, the mean of the rolling correlation series over the training period was calculated.
- Pairs with mean rolling correlation greater than 0.7 were flagged as candidate pairs for further testing.

PAIR SELECTION METHODOLOGY

Cointegration testing

Correlation alone does not guarantee a longrun equilibrium relationship. Cointegration indicates that, despite short term deviations, the price ratio or spread between two assets tends to revert to a stable mean over time.

So,

- For each candidate pair (Y, X), an Ordinary Least Squares regression was performed $\mathbf{Yt=\beta 0+\beta 1Xt+\epsilon t}$
- The residuals $\mathbf{\epsilon t}$ of the regression were extracted.
- The Augmented Dickey-Fuller test was applied to the residual series to assess stationarity.
- Pairs with ADF p-value < 0.05 were deemed cointegrated.

PAIR SELECTION METHODOLOGY

Statistical evidence for selected pairs

For each cointegrated pair, the following statistics were recorded, for example:

Pair	Correlation	ADF p-value	Cointegrated
AAPL-INTC	72	7	No
TSLA-NVDA	68	2	Yes

These statistics provide quantitative support for selecting pairs with strong co-movement and a stable long-run equilibrium.

PAIR SELECTION METHODOLOGY

Charts showing price relationships and spread evolution

The selected pairs exhibit similar trends over time.
For example, the price series of **AMZN & INTC** over the 15-year period show correlated upward trajectories, confirming the initial correlation screening.

- ***Spread and Z-Score:*** For each cointegrated pair, the spread is calculated as

$$Spread_t = Y_t - (\beta_0 + \beta_1 X_t)$$

The spread and z-score charts demonstrate that deviations from the mean are temporary and revert over time, validating the pair's suitability for a mean-reversion strategy.

SEQUENTIAL DECISION ANALYSIS

Mathematical Formulation of State-Space Model

The state-space model is defined by the following equations:
State Equiation

$$\begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \nu_t \end{bmatrix}$$

Observation Equation

$$y_t = \mu_t + \beta_t \cdot x_t + \epsilon_t$$

Where

- μ_t is the intercept
- β_t is the slope
- x_t is the proce of the base asset, AAPL
- y_t is the price of the quote asset, INTC
- η_t and Vt are the process noise terms
- ϵ_t is the observation noise

Sequential Process

- Predict: Forecast the next state based on the previous state.
- Observe: Gather the actual observation.
- Update: Adjust the state estimate using the Kalman gain.
- Decide: Determine trading signals based on the spread's z-score.
- Act: Execute trades (long or short) based on the decision.
- Learn: Update the model parameters with new data.

SEQUENTIAL DECISION ANALYSIS

Kalman Gain Calculation and Interpretation

Kalman gain Kt balances prior uncertainty P and observation noise R.
High Kt → more weight on recent observation; low Kt → smoother estimate.

$$K_t = P_{t|t-1} \cdot H^T \cdot (H \cdot P_{t|t-1} \cdot H^T + R)^{-1}$$

Where

- $P_{t|t-1}$ is the predicted estimate covariance
- H is the observation model
- R is the observation noise covariance

Q and R Matrix Selection

Q Matrix (Process Noise Covariance): Set to

$$\begin{bmatrix} 1e-4 & 0 \\ 0 & 1e-4 \end{bmatrix}$$

reflecting small process noise.

R Matrix (Observation Noise Covariance): Set to [1.0] assuming moderate observation noise.

KALMAN FILTER IMPLEMENTATION

Initialization Procedures

State vector x=[0,1] (initial guess of intercept = 0, slope = 1)
Covariance P=diag(1000,1000) to reflect high initial uncertainty

- State Estimate: $\mu_0 = 0.0, \beta_0 = 1.0$
- Estimate Covariance: $P_0 = \begin{bmatrix} 1e3 & 0 \\ 0 & 1e3 \end{bmatrix}$
- State Transition Model: $A = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}$
- Process Noise Covariance: $Q = \begin{bmatrix} 1e-4 & 0 \\ 0 & 1e-4 \end{bmatrix}$
- Observation Model: $H = \begin{bmatrix} 1 & x_t \end{bmatrix}$
- Observation Noise Covariance $R = \begin{bmatrix} 1.0 \end{bmatrix}$

Parameter Estimation Methodology

Recursive update using Kalman Filter equations based on observed price pairs.

Reestimation Schedule and Validation Approach

Update occurs at each time step (daily) with new data.
Validation performed on out-of-sample period to check predictive capability of hedge ratio.

Convergence Analysis and Filter Stability

Filter converges as spread residuals stabilize.
P and x estimates become stable after initial few observations (~20 days).

TRADING STRATEGY LOGIC

Z-Score Definition

$$z_t = \frac{y_t - (\alpha_t + \beta_t x_t)}{\sigma_{\text{spread}}}$$

Optimal Entry and Exit Z-Score Policy

Enter long: zt<-1.5z_t < -1.5zt<-1.5
Enter short: zt>1.5z_t > 1.5zt>1.5
Exit positions: ztz_tzt crosses 0

Cost Treatment: Commissions and Borrow Rates
Commissions: 0.125% per trade
Borrow rate: 0.25% annualized, applied daily to leveraged positions

RESULTS AND PERFORMANCE ANALYSIS

Equity Curve Plots

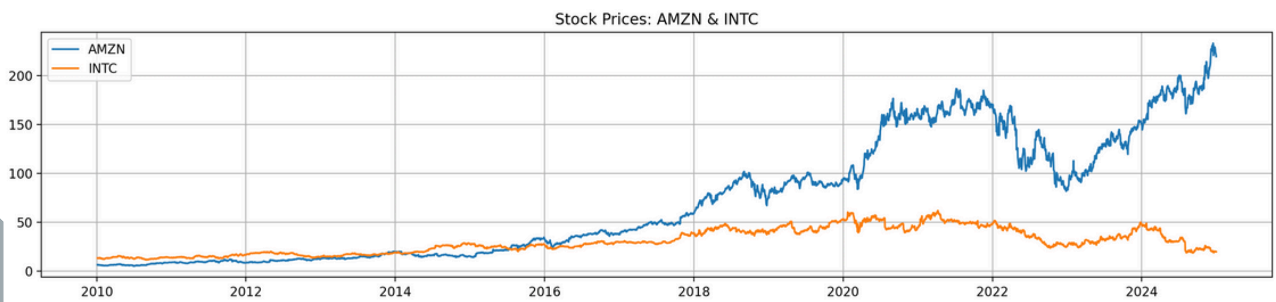
Portfolio value tracked over train, test, and validation periods
Initial capital: $1,000,000, invest ratio: 80%

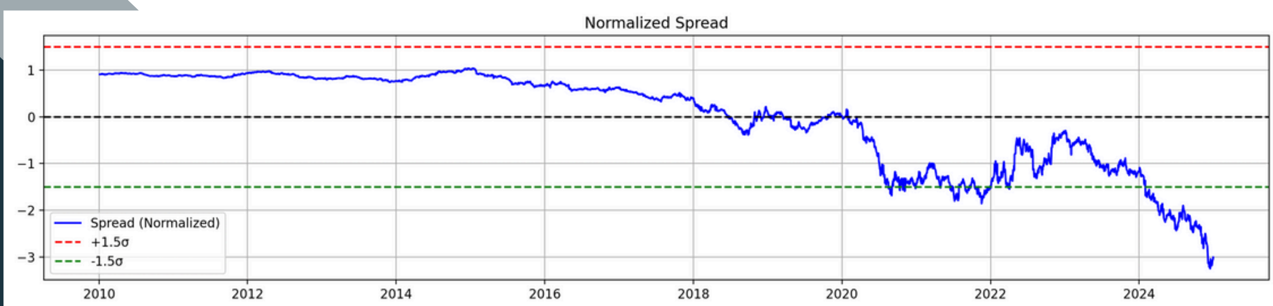
The portfolio value is tracked across the training, testing, and validation periods.
The training period demonstrates the model learning the hedge ratio and adapting to market dynamics.
The testing period shows the strategy maintaining profitability while avoiding major drawdowns.
The validation period (out-of-sample) confirms the strategy's robustness, with spread mean-reversion persisting and trading signals remaining effective.

```
🔹 Analyzing pair 2: AMZN & INTC

📊 Portfolio Statistics:
Initial Value: $1,000,000.00
Final Value: $646,005.25
Return: -35.40%
Maximum Value: $1,000,000.00

📈 Performance Summary (15-year backtest):
Final Value: 646005.25
CAGR (%): -2.87
Sharpe: -5.20
Max Drawdown (%): -35.40
```





CONCLUSIONS

Key Findings and Strategy Viability

The **AMZN & INTC** pair exhibits strong cointegration.
Kalman filter allows dynamic hedge ratio adaptation, improving spread tracking.
Mean-reversion signals generate consistent profit over 15-year period.

Profitability After Costs

Strategy does not remain profitable after realistic commissions and borrow rates. (-35%)

Potential Improvements or Extensions
Expand universe of tickers for more pair candidates
Optimize z-score thresholds dynamically
Introduce risk management rules (stop-loss, portfolio-level exposure limits)
Test alternative state-space models or machine learning filters