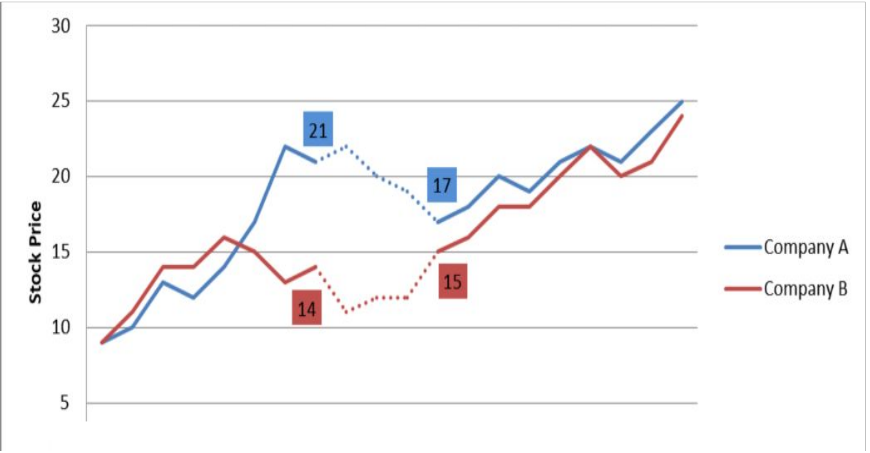
Focus of the project

•A pairs trade strategy is based on the historical correlation of two securities; the securities in a pairs trade must have a high positive correlation, which is the primary driver behind the strategy's profits.

•A pairs trade is a trading strategy that involves matching a long position with a short position in two stocks with a high correlation.



11%

Buy laggard, short leader; profit when the spread converges

Idea

30%+

Twin stocks deviate ~4–12% on average, sometimes ~40%—tradable gaps.

Important

β ≈ 0

Self-financing long–short spread; low market-direction risk, easy to replicate.

Chosen

Literature Review

Study	Methodology	Key Findings
Gatev, Goetzmann, Rouwenhorst (2006) *Pairs Trading: Performance of a Relative Value Arbitrage Rule*	Match pairs by minimizing the sum of squared deviations between the two normalized price series Trade on ±2σ divergence	13% annual excess return on U.S. equities (1962-2002) Profits robust after costs Capture temporal variation in returns different from simple mean reversion
De Jong, Rosenthal, Van Dijk (2009) The *Risk and Return of Arbitrage in Dual-Listed Companies*	Long-short arbitrage across 12 dual-listed companies (1980–2002) Exploiting price deviations from theoretical parity	10% abnormal return but high idiosyncratic volatility Show limits to arbitrage and horizon risk
Caldeira, Moura (2013) *Selection of a Portfolio of Pairs Based on Cointegration: A statistical Arbitrage Strategy*	Cointegration approach to form portfolios with best in-sample Sharpe ratios on the Brazilian market (2005-2012) Trade on spread z-score ±2σ thresholds	16% annual return Cointegration captures long-run equilibrium Strong even in crises
Chen, Wang, Sriboonchitta, Lee (2017) *Pair Trading based on Quantile Forecasting of Smooth Transition GARCH Models*	Minimum square distance pair selection Quantile forecast trading signals obtained from nonlinear smooth transition GARCH model	35.5% annualized return (18% with transaction costs) on U.S. stocks (2006-2014) Demonstrates volatility-aware strategy outperforms linear models

Data and Methods

Data:

- Source: Yahoo Finance for prototyping, WRDS/CRSP
- Log prices or normalized dividend adjusted prices over formation period (12M)
- Winsorize

Methods for pairs selection:

- Correlation

$$r_s = \rho_{rg_A, rg_B} = \frac{cov(rg_A, rg_B)}{\sigma_{rg_A} * \sigma_{rg_B}}$$

- Cointegration
  - Engle-Granger (Augmented Dickey-Fuller on residuals)

    $$P_t^A - \gamma P_B^t = \mu + \epsilon_t,$$

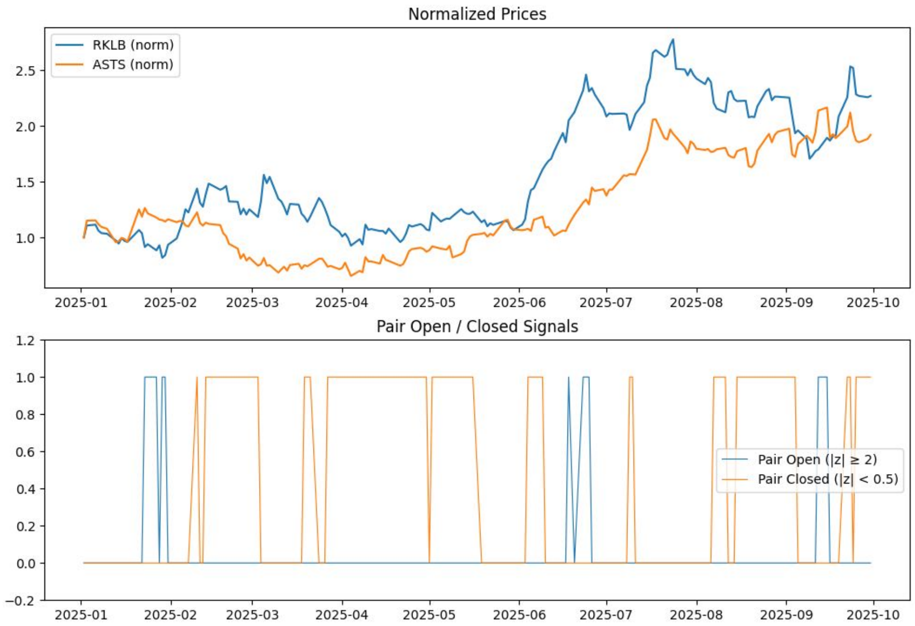
  - Johansen for small basket

- Minimum distance between normalized price (Gatev et al)

$$ESD = \sum_t (S_A(t) - S_B(t))^2,$$

Results: What do you predict your results to be? Why?

- Sample: Space tech stock



Normalized Prices: Pt(norm) = Pt / pt0

Pair open: when spread is wider

Pair closed: when spread is tighter

Z-score based on 2 sigma

Selection based on different industries: Even if two assets are from different sectors, if they have a **persistent equilibrium relationship**, they're fair game sharing for shared macro / factor exposure if statistically cointegrated ex) Tesla vs Carbon Credit ETF (Automotives vs ESG)

Risk and Backtesting

Backtesting Plan

- **Formation period:** 12 months; **Trading period:** 6 months
- **Entry/Exit rule:** Trade when Z-score of spread > ±2σ; close at mean reversion
- **Transaction cost:** 10–20 bps per trade
- **Metrics:** Return, Information ratio, hit rate, drawdown

Key Risks

- **Model risk:** Cointegration may not persist out-of-sample → use rolling window
- **Execution risk:** High turnover & illiquidity → apply liquidity filter
- **Regime risk:** Correlation breaks in crises → add stop-loss and recalibration
- **Data risk:** Missing or survivorship bias → use WRDS/CRSP clean data
- **Overfitting:** Excess tuning inflates results → use walk-forward validation