Project 09 — Pairs Trading

Cointegration + Kalman filter dynamic hedge ratio

Quant Finance Portfolio (Karim)

December 19, 2025

Abstract

This report supports **Project 09**, a market-neutral **pairs trading** workflow. It combines (i) **cointegration testing** to detect a stable long-run relationship between two assets, and (ii) a **Kalman filter** to estimate a **time-varying hedge ratio**, producing a dynamic spread. The project then builds a trading rule on the spread (z-score entry/exit), and evaluates performance with transaction costs and risk controls. The focus is on research discipline: correct statistical assumptions, clean signal construction, and robust backtesting without look-ahead.

Contents

1 What you build in Project 09

Pairs trading is a classical statistical arbitrage strategy: go long one asset and short the other when their **spread** deviates from a mean-reverting equilibrium.

This project implements an end-to-end pipeline:

- Download/clean price series for two assets and align calendars.
- Test for **cointegration** (Engle–Granger style) to justify a mean-reverting spread.
- Estimate a **static hedge ratio** (OLS) as a baseline.
- Estimate a **dynamic hedge ratio** using a **Kalman filter** (state-space model).
- Build the spread, compute a rolling **z-score**, and generate entry/exit signals.
- Backtest a **market-neutral** long/short strategy with transaction costs and risk metrics.
- Provide interactive visualizations: spread vs time, hedge ratio dynamics, trade markers, performance.

This project demonstrates applied time-series statistics, state-space modeling, and realistic backtesting for a market-neutral strategy.

2 Prerequisites (math and time-series foundations)

2.1 Returns, log-prices, and stationarity

Pairs trading is usually built on log-prices:

$$x_t = \log P_t^{(1)}, \qquad y_t = \log P_t^{(2)}.$$

A key distinction:

- **I(1)**: a series that becomes stationary after first differencing (common for prices),
- **I(0)**: stationary series (common for spreads / returns).

2.2 Regression and residuals (OLS)

Given y_t and x_t, the OLS regression

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

estimates a hedge ratio β. The residual ε_t is the candidate spread.

2.3 Cointegration (Engle–Granger intuition)

Two I(1) series (x_t, y_t) are cointegrated if there exists β such that the linear combination

$$s_t = y_t - \beta x_t$$

is stationary (I(0)). This supports the economic idea of a long-run equilibrium and justifies mean-reversion trades.

2.4 Kalman filter (state-space model intuition)

The Kalman filter estimates latent states in linear Gaussian models. Here, the latent state is a time-varying hedge ratio:

$$\beta_t \text{ evolves slowly over time.}$$

This captures changing relationships (regimes) between the two assets.

3 Cointegration and spread construction

3.1 Static hedge ratio (baseline)

Compute OLS on a training window to estimate $(\hat{\alpha}, \hat{\beta})$ and define the static spread:

$$s_t^{\mathrm{OLS}} = y_t - \hat{\alpha} - \hat{\beta} x_t.$$

Then test stationarity of s_t^{OLS} (ADF test), which is the practical Engle–Granger step.

3.2 Interpreting the cointegration test

If the residual spread is stationary, we treat it as mean-reverting. If not, a pairs strategy based on mean reversion is statistically unsupported (at least in that window).

4 Dynamic hedge ratio with a Kalman filter

4.1 State-space model

A standard formulation:

$$\begin{aligned} y_t &= \alpha_t + \beta_t x_t + \varepsilon_t, \qquad && \varepsilon_t \sim \mathcal{N}(0, R), \\ \begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} &= \begin{pmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{pmatrix} + \eta_t, \qquad && \eta_t \sim \mathcal{N}(0, Q). \end{aligned}$$

The state noise Q controls how quickly the hedge ratio can move (adaptivity).

4.2 Filtered spread

Using the filtered estimates $(\hat{\alpha}_t, \hat{\beta}_t)$, define the dynamic spread:

$$s_t^{\mathrm{KF}} = y_t - \hat{\alpha}_t - \hat{\beta}_t x_t.$$

This spread can be more stable in the presence of regime changes.

5 Trading rule: z-score entry/exit

Given a spread series s_t, define a rolling mean and standard deviation:

$$\mu_t = \mathrm{MA}_W(s)_t, \qquad \sigma_t = \mathrm{Std}_W(s)_t,$$

and the z-score:

$$z_t = \frac{s_t - \mu_t}{\sigma_t}.$$

A typical rule:

- enter a position when $|z_t| > z_{\text{entry}}$,
- exit when $|z_t| < z_{\text{exit}}$,
- optional stop if $|z_t|$ explodes (breakdown regime).

5.1 Position construction (market neutral)

A simple market-neutral position for the dynamic model:

$$\text{Long/short exposure:} \quad \begin{cases} \text{Short } y, \text{ long } \beta_t x & \text{if } z_t > z_{\text{entry}} \\ \text{Long } y, \text{ short } \beta_t x & \text{if } z_t < -z_{\text{entry}}. \end{cases}$$

In implementation, weights are normalized to control gross exposure and leverage.

6 Backtesting and evaluation

6.1 No look-ahead discipline

All rolling statistics (mean/vol of spread), hedge ratios, and signals are computed using information available up to time t, and trades are executed with a one-step lag when required by the data frequency.

6.2 Transaction costs and turnover

Pairs strategies can trade frequently. A transparent cost model applies a cost proportional to turnover:

$$R_t^{\text{net}} = R_t^{\text{gross}} - c \cdot \text{Turnover}_t.$$

This prevents unrealistic performance from excessive signal flipping.

6.3 Key reported metrics

- total return and annualized return,
- annualized volatility and Sharpe ratio,
- maximum drawdown,
- trade count, average holding period, turnover,
- cost drag (gross vs net performance).

7 Robustness checks and pitfalls

7.1 What makes a pairs backtest credible

- verify cointegration on a training window, then evaluate out-of-sample,
- test sensitivity to rolling window W and thresholds $(z_{\text{entry}}, z_{\text{exit}})$,
- compare static OLS vs dynamic Kalman hedge ratio,
- verify the strategy is not just a disguised market beta (check net exposure).

7.2 Common pitfalls

- **Spurious regression:** OLS on non-stationary prices can look significant without cointegration.
- **Regime breaks:** cointegration can disappear; the spread stops mean-reverting.
- **Overfitting:** tuning thresholds on the same sample inflates performance.
- **Hidden beta:** poor sizing can create unintended directional exposure.

8 Implementation notes (what the notebook is doing)

- Align and clean two asset price series; compute log-prices.
- Fit OLS and compute an OLS spread; run stationarity/ADF check.
- Run Kalman filter to estimate (α_t, β_t) and compute the dynamic spread.
- Build z-score signals and execute a market-neutral backtest with costs.
- Export interactive visuals (spread, hedge ratio, trades, equity curve) as HTML for the repo.

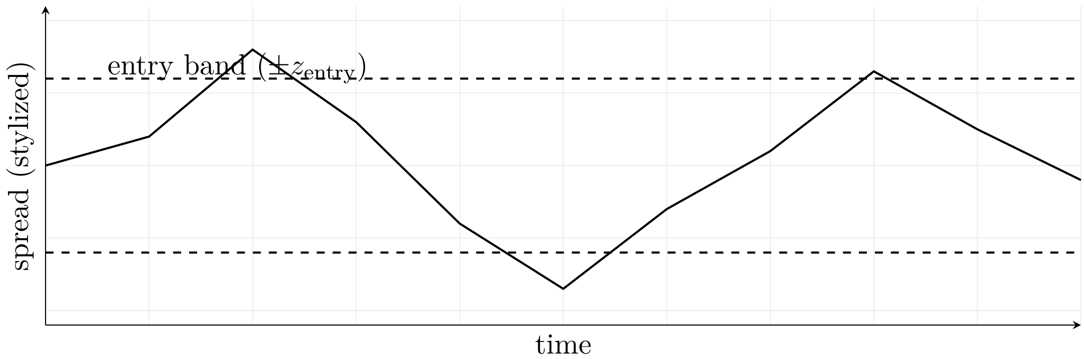
9 Sanity checks you should always do

- If the spread is not stationary (ADF fails), do not interpret deviations as mean reversion.
- Static and dynamic spreads should be on comparable scales; verify normalization.

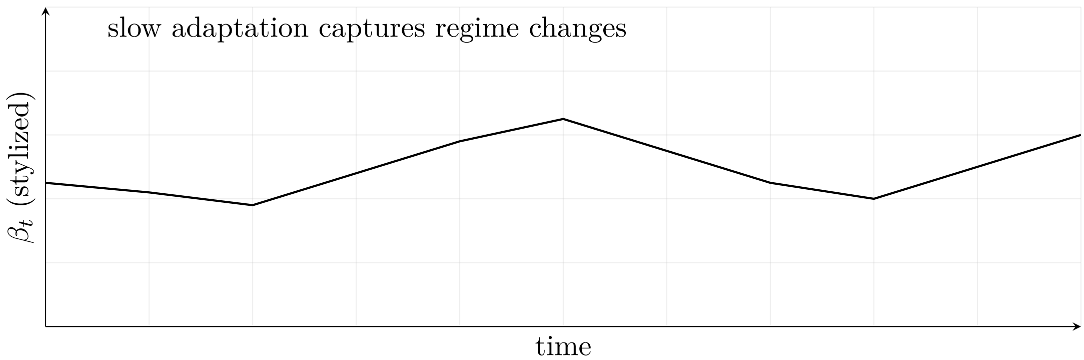
- Increasing transaction costs should reduce performance; high turnover should be penalized.
- Check that the strategy has near-zero average market exposure (market-neutral by construction).
- Verify no look-ahead: shifting signals by one day should change results materially.

10 Overleaf plots (conceptual, fast to compile)

10.1 Spread and z-score intuition (stylized)



10.2 Dynamic hedge ratio (stylized)



11 Interview pitch

This project implements an end-to-end market-neutral pairs trading workflow. I test for cointegration to justify a mean-reverting spread, then estimate both a static OLS hedge ratio and a time-varying hedge ratio using a Kalman filter (state-space model). I trade the resulting spread with a z-score entry/exit rule, control leverage and turnover, include transaction costs, and report professional metrics (Sharpe, drawdown, turnover, cost drag) with interactive diagnostics (spread, hedge ratio, trade markers).