Portfolio Optimization

Pairs Trading Portfolios

Daniel P. Palomar (2025). *Portfolio Optimization: Theory and Application.*
Cambridge University Press.

portfoliooptimizationbook.com

Latest update: 2025-11-13

Outline

Executive Summary

- **Strategy Origins**: Relative-value arbitrage approach known since the mid-1980s.
- **Core Concept**: Identifies two securities with historically correlated prices; trades the divergence by buying undervalued and shorting overvalued assets.
- **Profit Mechanism**: Realizes gains when prices revert to historical equilibrium (contrarian approach).
- **Mathematical Foundation**: Creates a mean-reverting virtual asset that exploits relative mispricings while maintaining market neutrality.
- **Beyond Pairs**: Extends to statistical arbitrage using multiple assets.
- **Scope of Presentation**: Introduction to fundamental concepts, pair discovery process, and advanced Kalman modeling techniques for trading. For details, see (Palomar 2025, chap. 15).

Outline

Introduction

Mean-Reversion:

- Property of a time series indicating a tendency to revert to a long-term average value.
- Crucial for pairs trading, where a mean-reverting series is constructed from two or more assets.
- Traders buy at low prices expecting a return to the long-term mean, closing positions for profit when prices revert.
- Assumes historical relationships between assets will persist, requiring careful monitoring.

Types of Mean-Reversion:

- **Longitudinal or time series mean-reversion:**
  Occurs along the time axis with a long-term average value.
  Deviations occur at different times in opposite directions.
- **Cross-sectional mean-reversion:**
  Occurs along the asset axis with an average value across assets.
  Some assets deviate in one direction, others in the opposite direction.

Introduction

Stationarity:

- Related but different from mean-reversion.
- Refers to fixed statistics of a time series over time.
- A stationary time series can be mean-reverting, but not all mean-reverting series are stationary.

Unit-Root Stationarity:

- Specific type of stationarity modeled with autoregressive (AR) model without unit roots.
- A time series with a unit root is non-stationary and tends to diverge over time.
- Example of unit-root nonstationarity: random walk model for log-prices:

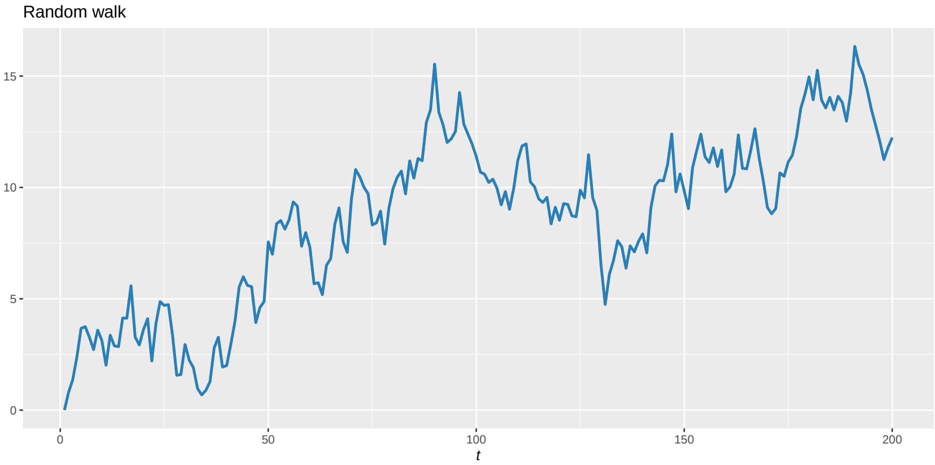
$$y_t = \mu + y_{t-1} + \epsilon_t,$$

  where μ is the drift and ϵ_t the residual.
- Example of AR(1) model without unit root (mean reversion):

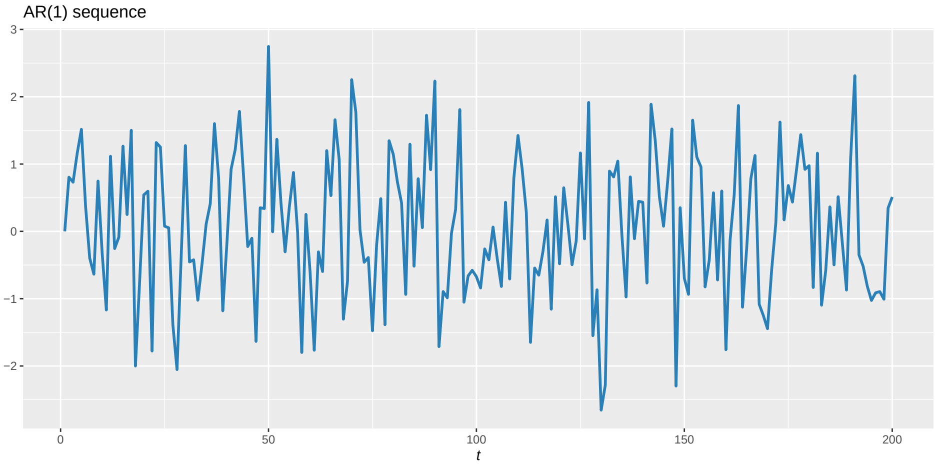
$$y_t = \mu + \rho\, y_{t-1} + \epsilon_t,$$

  where $|\rho| < 1$.

Example of a random walk (nonstationary time series with unit root):



Example of a unit-root stationary AR(1) sequence:



Practical Implications:

- Mean-reversion and unit-root stationarity are not equivalent but unit-root stationarity is a practical proxy for mean-reversion.
- Testing for unit-root stationarity is the standard approach for determining mean-reversion.

Differencing:

- Operation used to obtain stationarity.
- Involves taking differences between consecutive samples: $\Delta y_t = y_t - y_{t-1}$.
- Can make a nonstationary time series, like a random walk, become stationary.
- Example: differencing log-prices to obtain log-returns, indicating log-prices are integrated of order 1.

Cointegration

Cointegration:

- Property where two (or more) assets, while not mean-reverting individually, are mean-reverting with respect to each other.
- Occurs when series contain stochastic trends (nonstationary) but move closely together, making their difference stable (stationary).
- Mimics a long-run equilibrium in an economic system.

Intuitive Example:

- Drunken man and dog wandering the streets: both paths are nonstationary, but the distance between them is mean-reverting and stationary.

Mathematical Definition:

- A multivariate time series $\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, \ldots$ is cointegrated if some linear combination becomes integrated of lower order.
- If $\boldsymbol{y}_t$ is nonstationary but $\boldsymbol{w}^\mathsf{T}\boldsymbol{y}_t$ is stationary for some weights $\boldsymbol{w}$.
- Example: log-prices of stocks are nonstationary, but a linear combination $\boldsymbol{w}^\mathsf{T}\boldsymbol{y}_t$ can be stationary.

Cointegration

Random walk by a drunken man with a dog:

Cointegration

Modeling Cointegration: Common model for two time series:

$$y_{1t} = \gamma\, x_t + w_{1t}$$
$$y_{2t} = x_t + w_{2t},$$

where

- x_t is a stochastic common trend defined as a random walk:

$$x_t = x_{t-1} + w_t,$$

- w_{1t}, w_{2t}, w_t are i.i.d. residual terms, with variances σ_1^2, σ_2^2, and σ^2,
- γ is the coefficient determining the cointegration relationship.

Implications:

- Each time series y_{1t} and y_{2t} is a random walk plus noise, hence nonstationary.
- Sharing a common stochastic trend allows a linear combination to eliminate this trend.
- The **spread** is the linear combination without the trend:

$$z_t = y_{1t} - \gamma\, y_{2t} = w_{1t} - \gamma\, w_{2t}$$

- The spread z_t is stationary and mean-reverting.

Correlation

Correlation:

- Basic concept in probability measuring how correlated two random variables are.
- Applicable to stationary time series, not nonstationary time series.
- In finance, correlation is used for asset returns, not price values.

Calculating Correlation:

- Given two time series of log-prices, y_{1t} and y_{2t}, obtain log-returns as differences Δy_{1t} and Δy_{2t}.
- Correlation defined assuming stationarity:

$$\rho = \frac{\mathbb{E}\left[(\Delta y_{1t} - \mu_1) \cdot (\Delta y_{2t} - \mu_2)\right]}{\sqrt{\mathrm{Var}(\Delta y_{1t}) \cdot \mathrm{Var}(\Delta y_{2t})}},$$

  where μ_1 and μ_2 are the means of Δy_{1t} and Δy_{2t}.
- Correlation is bounded: $-1 \leq \rho \leq 1$ (thanks to the normalization in the denominator).

Interpretation of Correlation:

- High correlation: two time series co-move (move simultaneously in the same direction).
- Zero correlation: two time series move independently.

Correlation vs. Cointegration

Correlation vs. Cointegration:
Both concepts capture the similarity of movements of two time series but are fundamentally different in their definitions and implications.

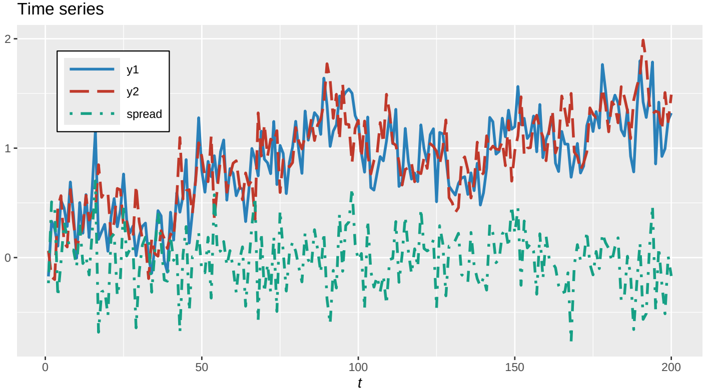
Analytical Derivation:

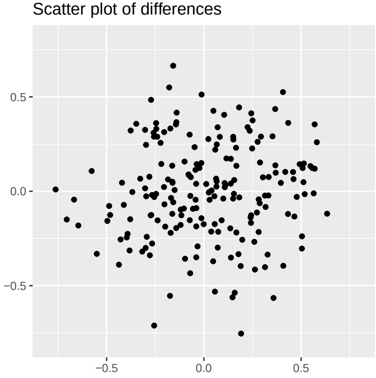
- For the cointegrated time series with the stochastic common trend model:

$$\rho = \frac{1}{\sqrt{1 + 2\frac{\sigma_1^2}{\sigma^2}}\sqrt{1 + 2\frac{\sigma_2^2}{\sigma^2}}},$$

- Correlation can be made arbitrarily small by choosing appropriate variances σ_1^2, σ_2^2, and σ^2.
- Perfectly cointegrated time series can have very low correlation, highlighting the difference between the two concepts.

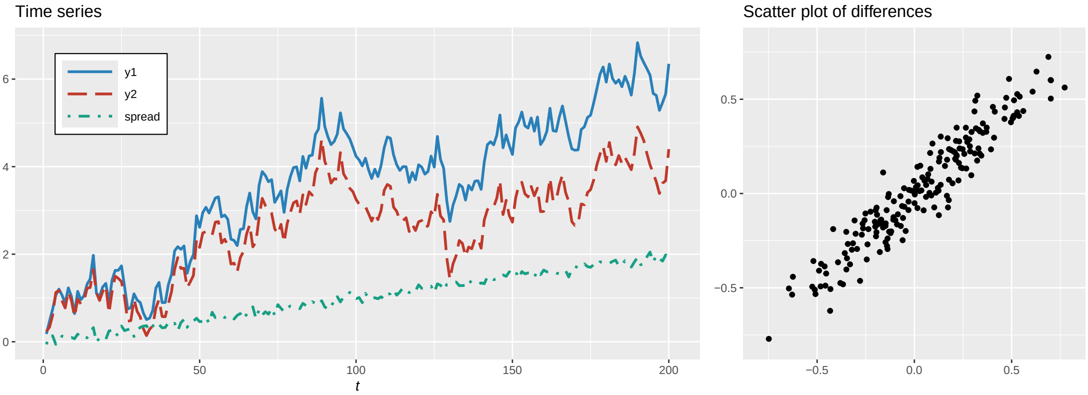
Example of cointegrated time series with low correlation:





Correlation vs. Cointegration

Example of non-cointegrated time series with high correlation:



Correlation vs. Cointegration

Key Differences:

- **Correlation:**
  - High when two time series co-move (move simultaneously in the same direction).
  - Zero when they move independently.
- **Cointegration:**
  - High when two time series move together and remain close to each other.
  - Nonexistent when they do not stay together.

Short-Term vs. Long-Term:

- **Correlation:**
  - Concerned with short-term movements (directional movement from one period to the next).
  - Ignores long-term trends.
- **Cointegration:**
  - High when two time series move together and remain close to each other.
  - Nonexistent when they do not stay together.

Correlation vs. Cointegration

Define the difference of a time series y_t over k periods as $r_t(k) = y_t - y_{t-k}$ for $t = 0, \ldots, T$. (Note that $r_t(1)$ is just the first difference Δy_t.)

Precise Interpretation:

- **Correlation:** uses 1-period differences $r_{1t}(1) = \Delta y_{1t}$ and $r_{2t}(1) = \Delta y_{2t}$.
- **Cointegration:** uses t-period differences $r_{1t}(t) = y_{1t} - y_{10}$ and $r_{2t}(t) = y_{2t} - y_{20}$.

Pairs Trading:

- Cointegration is crucial, not correlation.
- Focuses on the long-term mean-reversion property.

Outline

Pairs Trading

Historical Context:

- Developed in the mid-1980s by Nunzio Tartaglia's team at Morgan Stanley.
- Achieved significant success but the team disbanded in 1989.
- Technique spread across the quant community after the initial secrecy was lost.

Trading Strategies Classification:

- **Momentum-based strategies (or directional trading):**
  Capture market trends.
  Treat fluctuations as undesired noise (risk).
- **Pairs trading (or statistical arbitrage):**
  Market neutral.
  Trade mean-reverting fluctuations of relative mispricings between two securities.

Pairs Trading

Mean-Reversion Trading:

- Buy when the asset is below its mean value and sell when it recovers.
- Short-sell when the asset is above its mean value and buy back when it reverts.
- Directly finding a mean-reverting asset in financial markets is virtually impossible.

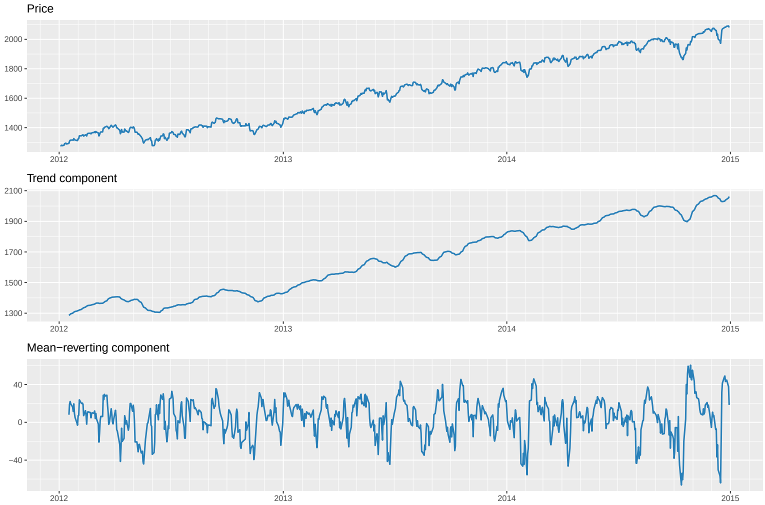
Creating a Mean-Reverting Asset:

- Discover a cointegrated pair of assets.
- Create a virtual mean-reverting asset (spread) from the pair.
- The spread is market neutral, as it does not follow the market trend.

Pairs Trading:

- Market-neutral strategy trading a mean-reverting spread.
- Identifies two historically cointegrated financial instruments (e.g., stocks).
- Takes long and short positions when prices deviate from their historical mean relationship.
- Profits from the convergence back to the historical equilibrium.
- References: (Vidyamurthy 2004), (Ehrman 2006), (Chan 2013), (Feng and Palomar 2016).

Decomposition of asset price into trend component and mean-reverting component:



Spread

Simplest Implementation of Pairs Trading:

- Based on comparing the spread of two time series y_{1t} and y_{2t} to a threshold s_0.
- Spread defined as:

$$z_t = y_{1t} - \gamma\, y_{2t}$$

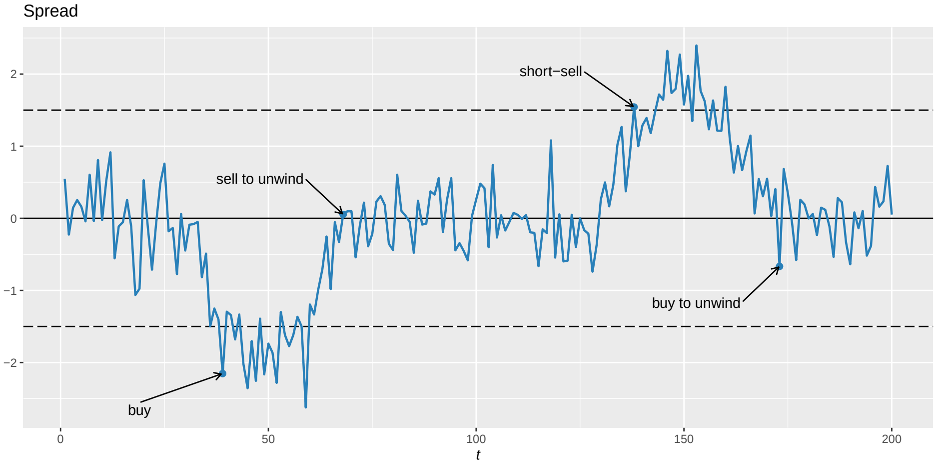
  which is assumed to be mean-reverting with mean μ.

Trading Strategy:

- **Buy signal:** buy if the spread is low: $z_t < \mu - s_0$.
- **Short-sell signal:** short-sell if the spread is high: $z_t > \mu + s_0$.
- **Unwinding the position:**
  Unwind the position when the spread reverts back to the mean (after k periods).
  This ensures a difference of at least $|z_{t+k} - z_t| \geq s_0$.

Spread

Illustration of pairs trading via thresholds on the spread:



Prices vs. Log-Prices

Pairs Trading Implementation:

- Can be implemented in terms of prices or log-prices.
- Determined by whether cointegration is exhibited by time series of prices or log-prices.
- Interpretation differs slightly based on this distinction.

Prices. If y_{1t} and y_{2t} represent the prices of two assets:

- Mean-reverting spread: $z_t = y_{1t} - \gamma\, y_{2t}$.
- Coefficients (1 and γ) represent the number of shares.
- Spread has the meaning of price value.
- Spread difference corresponds to profit over k periods (ignoring transaction costs):

$$z_{t+k} - z_t = s_0.$$

Prices vs. Log-Prices

Log-Prices. If y_{1t} and y_{2t} represent the log-prices of two assets:

- Use portfolio notation:
$$\boldsymbol{w} = \begin{bmatrix} 1 \\ -\gamma \end{bmatrix}$$
- Coefficients (1 and γ) represent normalized dollar values.
- Spread written as $z_t = y_{1t} - \gamma\, y_{2t} = \boldsymbol{w}^\mathsf{T} \boldsymbol{y}_t$, where $\boldsymbol{y}_t = \left[y_{1t}, y_{2t}\right]^\mathsf{T}$.
- Spread difference corresponds (approximately) to the return over k periods (ignoring transaction costs):
$$\boldsymbol{w}^\mathsf{T} \left(\boldsymbol{y}_{t+k} - \boldsymbol{y}_t\right) = z_{t+k} - z_t = s_0.$$

Summary:

- **Prices:**
    - Threshold s_0 determines absolute profit over k periods.
    - Number of shares stays constant, no rebalancing required.
- **Log-prices:**
    - Threshold s_0 determines the return over k periods.
    - Portfolio $\boldsymbol{w}$ represents normalized dollars, may require rebalancing.

Is Pairs Trading Profitable?

Assumptions and Risks in Pairs Trading:

- Relies on assumption that the historical relationship between two instruments will persist.
- Cointegration between financial instruments can change over time due to market conditions, industry trends, company-specific events, etc.
- Traders should carefully monitor the relationship and use risk management techniques.

Evidence and Publications:

- Some studies show pairs trading can provide profits (Elliott, Van Der Hoek, and Malcolm 2005), (Gatev, Goetzmann, and Rouwenhorst 2006), (Avellaneda and Lee 2010).
- Other studies indicate cointegration relationships may not be preserved over time (Chan 2013), (Clegg 2014).

Caution with Positive Results:

- Backtests may ignore transaction costs, which can exceed profits (Chan 2008).
- Strategies may have yielded past profits but may be less effective in recent times (Chan 2013).

Design of Pairs Trading

Overview of Pairs Trading:

- Relies on cointegrated pairs.
- Objective: trade profitably a mean-reverting spread.

Key Requirements:

1. **Discovering cointegrated pairs:** methods range from simple pre-screening to sophisticated statistical tests.
2. **Designing the trading strategy:** involves choosing the threshold s_0 or other sophisticated methods.

Advanced Topics:

- **Kalman filtering:** for estimating a time-varying cointegration relationship.
- **Extension to more than two assets:** statistical arbitrage.

Outline

Pre-Screening

Pre-Screening for Pairs Trading:

- Simple and cost-effective process to discard many pairs and select potential pairs for further analysis.

Normalized Price Distance (NPD):

- Common heuristic proxy for cointegration:

$$\mathsf{NPD} \triangleq \sum_{t=1}^{T} \left(\tilde{p}_{1t} - \tilde{p}_{2t}\right)^2$$

  where $\tilde{p}_{1t}$ and $\tilde{p}_{2t}$ are normalized prices ($\tilde{p}_{it} = p_{it}/p_{i0}$).
- If NPD large, discard the pair; otherwise keep for subsequent more serious cointegration tests.

Cointegration Tests

Cointegration Tests:

- Developed in statistics literature to check if a linear combination of two time series is stationary and mean-reverting.
- A time series with a unit root is nonstationary (random walk).
- Absence of unit roots indicates a tendency to revert to a long-term mean.
- Cointegration tests are typically implemented via unit-root stationarity tests.

Mathematical Objective:

- Determine if there exists a value of γ such that the spread:

$$z_t = y_{1t} - \gamma\, y_{2t}$$

  is stationary.
- Mean of the spread μ (equilibrium value) is not necessarily zero.
- γ does not have to be one; setting $\gamma = 1$ for dollar-neutral strategies reduces the number of cointegrated pairs.

Cointegration Tests

Heuristic Measures of Mean-Reversion:

- **Mean-crossing rate:** number of times the signal crosses its mean value over time (Vidyamurthy 2004).
- **Half-life of mean-reversion:** time it takes for a time series to return to within half the distance from the mean after deviating (Chan 2013).

Engle–Granger Test:

- Simple and direct method for testing cointegration (Engle and Granger 1987).
- Two-step process:
  1. Obtain γ via least squares regression.
  2. Test the residual for stationarity.
- Regression model:

$$y_{1t} - \gamma \, y_{2t} = \mu + r_t$$

- Residual r_t is checked for unit-root stationarity or mean-reversion.

Cointegration Tests

Statistical Tests for Stationarity:

- Dickey–Fuller (DF)
- Augmented Dickey–Fuller (ADF)
- Phillips–Perron (PP)
- Pantula, Gonzales-Farias and Fuller (PGFF)
- Elliott, Rothenberg and Stock DF-GLS (ERSD)
- Johansen's Trace Test (JOT)
- Schmidt and Phillips Rho (SPR)

R Packages for Cointegration Tests:

- `urca`
- `egcm`

Dickey–Fuller (DF) Test:

- Simplest model for the residual:

$$r_t = \rho\, r_{t-1} + \epsilon_t$$

- Null hypothesis: unit root present ($\rho = 1$).
- Alternative hypothesis: series is stationary ($|\rho| < 1$).
- Small *p*-value indicates strong stationarity (rejection of null hypothesis).

Cointegration Tests for More Than Two Time Series

Drawbacks of Engle–Granger Cointegration Test:

- Designed for two time series (assets).
- Regression step is sensitive to the ordering of the variables.
- Extending to more than two assets increases sensitivity to variable ordering.

Johansen's Test:

- Alternative method for cointegration testing (Johansen 1991, 1995).
- Based on multivariate time series modeling (Palomar 2025, chap. 4).

Johansen's Test Procedure:

- Fits a multivariate VECM (Vector Error Correction Model) for N assets.
- Key component: $N \times N$ matrix $\mathbf{\Pi}$ characterizing cointegration.
- Analyzes the rank of matrix $\mathbf{\Pi}$ to determine the number of different cointegration relationships present.

Are Cointegrated Pairs Persistent?

Persistence of Cointegration:

- Discovering a cointegrated pair and passing tests does not guarantee persistent profitability.
- Cointegration may not be stable over time.

Challenges in Practice:

- Cointegrated pairs found in historical data may lose cointegration in subsequent out-of-sample periods (Chan 2013).
- Factors affecting persistence: management decisions, competition, company-specific news, etc.

Empirical Evidence:

- Studies show that cointegration is not always a persistent property (Clegg 2014).
- Spread series of pairs are often affected by permanent shocks disrupting cointegration.

Addressing Practical Problems:

- **Time-varying cointegration:** via Kalman filter.
- **Relaxed forms of cointegration:** concept of partial cointegration allows the spread to contain a random walk component (Clegg and Krauss 2018).

Numerical Experiments: Synthetic Data Cointegrated

Example of Cointegrated but Uncorrelated Time Series:

- Previous example of synthetic cointegrated time series with low correlation.
- Based on $T = 200$ observations, the estimated cointegration relationship is

$$
\begin{aligned}
y_{2t} &= 0.80\, y_{1t} + 0.20 + r_t \\
r_t &= 0.12\, r_{t-1} + \epsilon_t
\end{aligned}
$$

- Residual r_t has a small autoregressive coefficient of 0.12, indicating no unit root.

Observations:

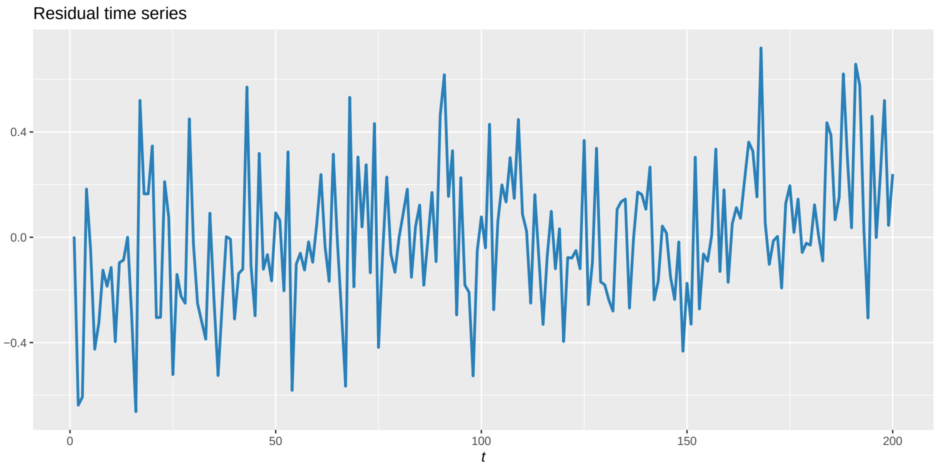
- Next figure plots the residual and shows strong mean-reversion.
- Estimated half-life of the residual is 0.33.

Quantitative Analysis:

- Next table provides p-values for several cointegration and residual unit-root tests.
- All p-values are below a reasonable threshold (e.g., 0.01).
- Null hypothesis (existence of a unit root) can be rejected.
- Cointegration of the two time series is accepted.

Numerical Experiments: Synthetic Data Cointegrated

Cointegration residual with cointegration and low correlation:



Numerical Experiments: Synthetic Data Cointegrated

Cointegration and residual unit-root tests:

Test	p-value
Augmented Dickey Fuller (ADF)	0.00805
Phillips-Perron (PP)	0.00010
Pantula, Gonzales-Farias and Fuller (PGFF)	0.00010
Elliott, Rothenberg and Stock DF-GLS (ERSD)	0.00081
Johansen's Trace Test (JOT)	0.00010
Schmidt and Phillips Rho (SPR)	0.00010

Numerical Experiments: Synthetic Data Non-Cointegrated

Example of Non-Cointegrated Time Series with High Correlation:

- Previous example of synthetic non-cointegrated time series with high correlation.
- Based on $T = 200$ observations, the estimated cointegration relationship is

$$y_{2t} = 0.68 y_{1t} + 0.16 + r_t$$
$$r_t = 0.91 r_{t-1} + \epsilon_t$$

- Residual r_t has a dangerous autoregressive coefficient of 0.91, which is close to 1, suggesting that the existence of a unit root cannot be excluded.

Observations:

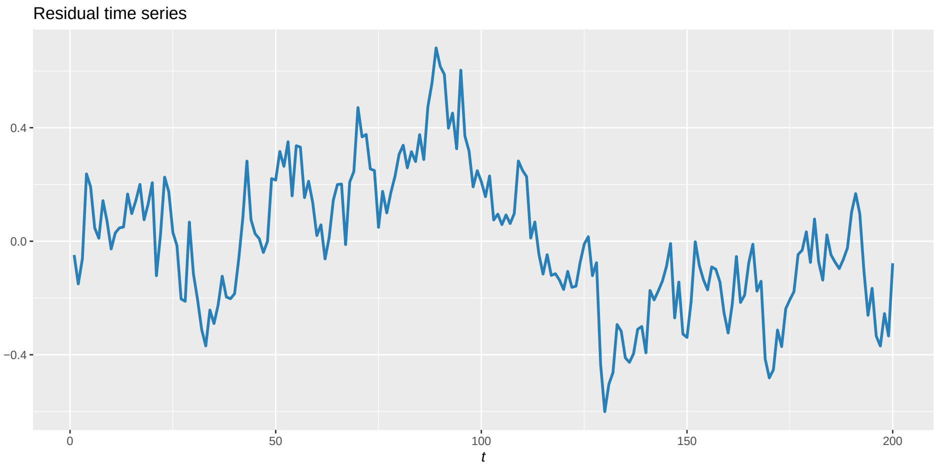
- Next figure plots the residual and corroborates the non-cointegration.
- Estimated half-life of the residual is 7.29 (weak mean-reversion).

Quantitative Analysis:

- Next table provides p-values for several cointegration and residual unit-root tests.
- All p-values are much higher than any reasonable threshold of, say, 0.01.
- Null hypothesis (existence of a unit root) cannot be rejected.
- Cointegration of the two time series cannot be concluded.

Numerical Experiments: Synthetic Data Non-Cointegrated

Cointegration residual with no cointegration and high correlation:



Numerical Experiments: Synthetic Data Non-Cointegrated

Cointegration and residual unit-root tests:

Test	p-value
Augmented Dickey Fuller (ADF)	0.4529
Phillips-Perron (PP)	0.0608
Pantula, Gonzales-Farias and Fuller (PGFF)	0.0700
Elliott, Rothenberg and Stock DF-GLS (ERSD)	0.0767
Johansen's Trace Test (JOT)	0.0996
Schmidt and Phillips Rho (SPR)	0.2671

Numerical Experiments: Market Data EWA–EWC

ETFs:

- **EWA:**
  ETF tracking the MSCI Australia Index.
  Includes Australian companies from sectors like financials, materials, healthcare, consumer staples, and energy.
- **EWC:**
  ETF tracking the MSCI Canada Index.
  Provides exposure to the Canadian equity market.
- Both ETFs offer broad exposure to the Australian and Canadian economies, respectively.

MSCI:

- Morgan Stanley Capital International (MSCI) is a leading provider of investment decision support tools and services.
- Known for its global equity indices used by investors to benchmark and analyze equity market performance.

Numerical Experiments: Market Data EWA–EWC

Cointegration of EWA and EWC:

- Popular example of cointegrated ETFs in the quant community (Chan 2013).
- Both economies are commodity-based, leading to related stock market performance through natural resources' prices.

Cointegration Relationship (2016-2019):

- Estimated via least squares:
  - **EWA regressed against EWC:** hedge ratio $\gamma = 0.74$.
  - **EWC regressed against EWA:** hedge ratio $\gamma = 1.27$ (different from $1/0.74 \approx 1.35$).
- **Johansen's test:** More accurate weights: 1 for EWA and -0.80 for EWC.

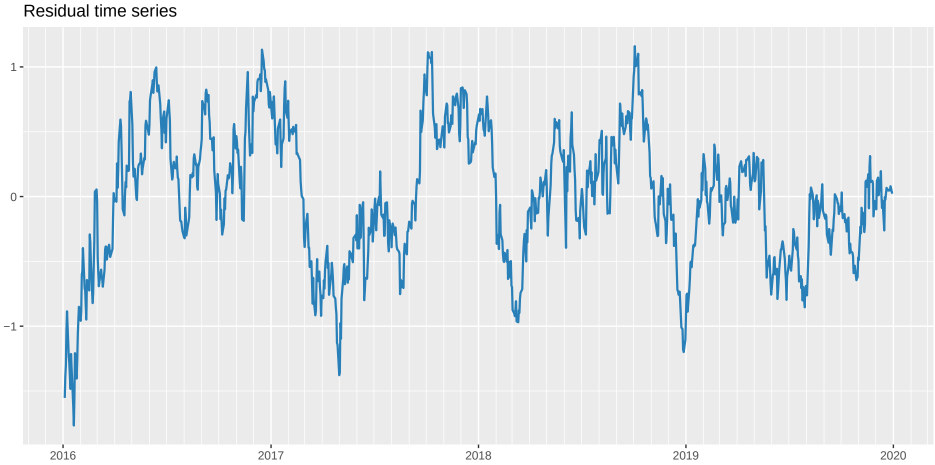
Residual Analysis:

- Next figure shows the residual of the cointegration relationship (spread).
- Estimated half-life of the residual: 19 days (indicating not very strong mean-reversion).

Cointegration Test Results:

- Next table shows the results for cointegration tests.
- Majority of tests indicate cointegration at the 1% level (p-value < 0.01).
- Two tests reject cointegration, suggesting caution should be taken.

Cointegration residual for EWA–EWC:



Numerical Experiments: Market Data EWA–EWC

Cointegration and residual unit-root tests for EWA–EWC:

Test	p-value
Augmented Dickey Fuller (ADF)	0.0049
Phillips-Perron (PP)	0.0058
Pantula, Gonzales-Farias and Fuller (PGFF)	0.0062
Elliott, Rothenberg and Stock DF-GLS (ERSD)	0.5310
Johansen's Trace Test (JOT)	0.0069
Schmidt and Phillips Rho (SPR)	0.3840

Numerical Experiments: Market Data KO–PEP

Coca-Cola (KO) and Pepsi (PEP):

- Often mentioned as a potential pair for pairs trading due to being in the same industry group.
- However, they do not seem to be cointegrated (Chan 2008).

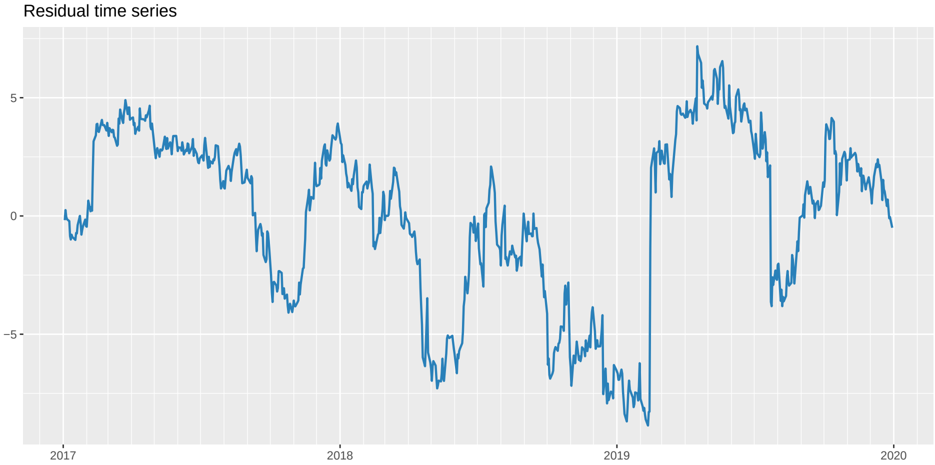
Cointegration Assessment (2017-2019):

- **Correlation:**
  Returns show a correlation of 0.66, which is statistically significant.
  Correlation is different from cointegration.
- **Least squares regression:**
  Used to assess the cointegration relationship.
- **Residual analysis:**
  Next figure shows the residual of the cointegration relationship (spread).
  Estimated half-life of the residual: 70 days (not indicative of cointegration).

Cointegration Test Results:

- Next table shows the results for cointegration tests.
- All tests reject the hypothesis of cointegration (all p-values are much larger than 0.01).

Cointegration residual for KO–PEP:



Cointegration and residual unit-root tests for KO–PEP:

Test	p-value
Augmented Dickey Fuller (ADF)	0.2675
Phillips-Perron (PP)	0.1845
Pantula, Gonzales-Farias and Fuller (PGFF)	0.1395
Elliott, Rothenberg and Stock DF-GLS (ERSD)	0.0484
Johansen's Trace Test (JOT)	0.5627
Schmidt and Phillips Rho (SPR)	0.1982

Numerical Experiments: Market Data SPY–IVV–VOO

SPY, IVV, and VOO:

- They are S&P 500 ETFs (i.e., track the S&P 500 index).
- Likely a strong cointegrating relationship due to tracking the same underlying asset.

Johansen's Test Results (2017-2019):

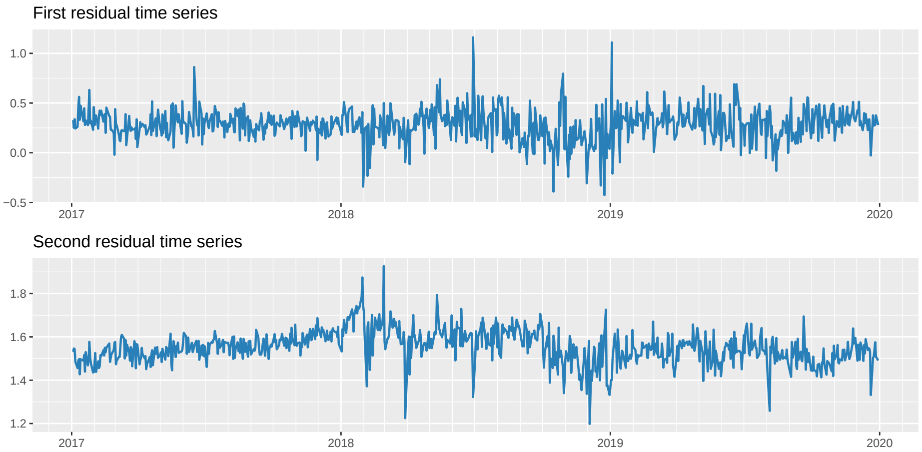
- Step 1: Null hypothesis: $r = 0$ vs. alternative hypothesis: $r > 0$.
  Clear evidence to reject the null hypothesis.
- Step 2: Null hypothesis: $r \leq 1$ vs. alternative hypothesis: $r > 1$.
  Sufficient evidence to reject the null hypothesis.
- Step 3: Null hypothesis: $r \leq 2$ vs. alternative hypothesis: $r > 2$.
  Cannot reject the null hypothesis.

Conclusion:

- Rank is $r = 2$.
- Two different cointegrating relationships can be found.
- Residuals of these cointegrating relationships are shown in next figure.

Numerical Experiments: Market Data SPY–IVV–VOO

Cointegration residuals for SPY–IVV–VOO:



Outline

Trading the Spread

Normalized Leverage for Cointegrated Pair:

- Given cointegrated log-price time series y_{1t} and y_{2t}.
- Form the spread $y_{1t} - \gamma\, y_{2t}$ using the 2-asset portfolio $\boldsymbol{w} = [1, -\gamma]^\mathsf{T}$.
- Leverage of the portfolio: $\|\boldsymbol{w}\|_1 = 1 + \gamma$.
- Normalize leverage to 1:

$$\boldsymbol{w} = \frac{1}{1+\gamma}\begin{bmatrix} 1 \\ -\gamma \end{bmatrix}.$$

- Corresponding normalized spread: $z_t = \boldsymbol{w}^\mathsf{T}\boldsymbol{y}_t$.

Portfolio Return:

- Return at time t (ignoring transaction costs):

$$\boldsymbol{w}^\mathsf{T}\left(\boldsymbol{y}_t - \boldsymbol{y}_{t-1}\right) = z_t - z_{t-1}$$

- Enter a position at time t and close after k periods when the spread reverts to the mean:

$$\left|z_{t+k} - z_t\right| \geq s_0$$

- This represents the portfolio return during these k periods.

Trading the Spread

Trading Signal:

- Decide when to buy or short-sell the spread and how much to invest (sizing).
- "Signal" time series $s_1, s_2, s_3, \dots$
- Signal s_t denotes the sizing (positive for buying, zero for no position, negative for short-selling).
- It is bounded as $-1 \leq s_t \leq 1$ to control leverage.
- Signal s_t is based on information up to (and including) time t.

Time-Varying Portfolio:

- Combination of the spread portfolio $\boldsymbol{w}$ and the signal s_t produces the time-varying portfolio $s_t \times \boldsymbol{w}$.
- Corresponding return:

$$R_t^{\text{portf}} = s_{t-1} \times \boldsymbol{w}^{\mathsf{T}} \left(\boldsymbol{y}_t - \boldsymbol{y}_{t-1}\right) = s_{t-1} \times \left(z_t - z_{t-1}\right).$$

Defining a Trading Strategy:

- Determine a rule for the sizing signal s_t.
- Use a normalized version of the spread, called the standard score or z-score:

$$z_t^{\text{score}} = \frac{z_t - \mathbb{E}[z_t]}{\sqrt{\mathrm{Var}(z_t)}},$$

- z-score has zero mean and unit variance but cannot be used directly (look-ahead bias).

Adaptive Calculation:

- Use training data to estimate mean and standard deviation, then apply to future data.
- More sophisticated approach: implement in a rolling fashion using Bollinger Bands.

Bollinger Bands:

- Created by John Bollinger in the early 1980s.
- Computed on a rolling window basis over a lookback window.
- Rolling mean and rolling standard deviation are used to obtain upper and lower bands (typically mean plus/minus one or two standard deviations).
- Adaptively normalize the spread with rolling mean and standard deviation.

Trading Strategies

Simple strategies for trading a spread: Contrarian nature of buying low and selling high.

- **Linear strategy:** (Chan 2013)
  - Define sizing signal as the negative z-score:

$$s_t = -\left[\frac{z_t^{\text{score}}}{s_0}\right]_{-1}^{+1}$$

  - s_0 denotes the threshold at which the signal is fully leveraged.
  - Project value to lie in the interval $[-1, 1]$ to limit leverage.
- **Thresholded strategy:** (Vidyamurthy 2004)
  - All-in or all-out sizing based on thresholds.
  - Compare z-score to a threshold s_0:

$$s_t = \begin{cases} +1 & \text{if } z_t^{\text{score}} < -s_0 \\ 0 & \text{after } z_t^{\text{score}} \text{ reverts to } 0 \\ -1 & \text{if } z_t^{\text{score}} > +s_0. \end{cases}$$

Illustration of pairs trading via the linear strategy on the spread:

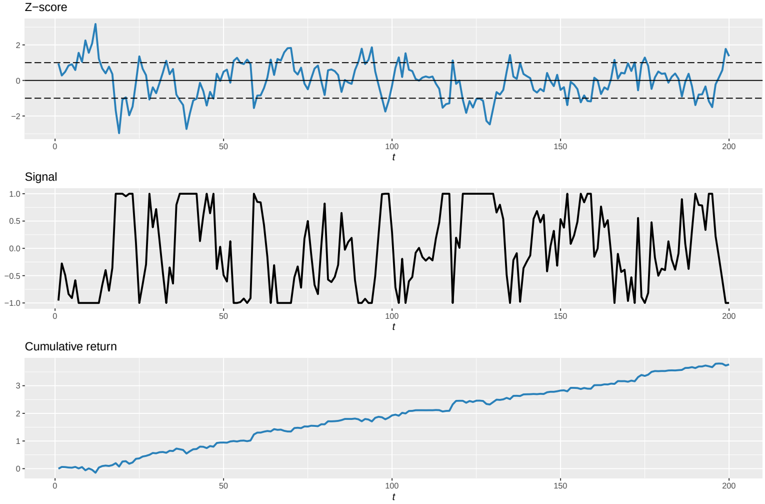
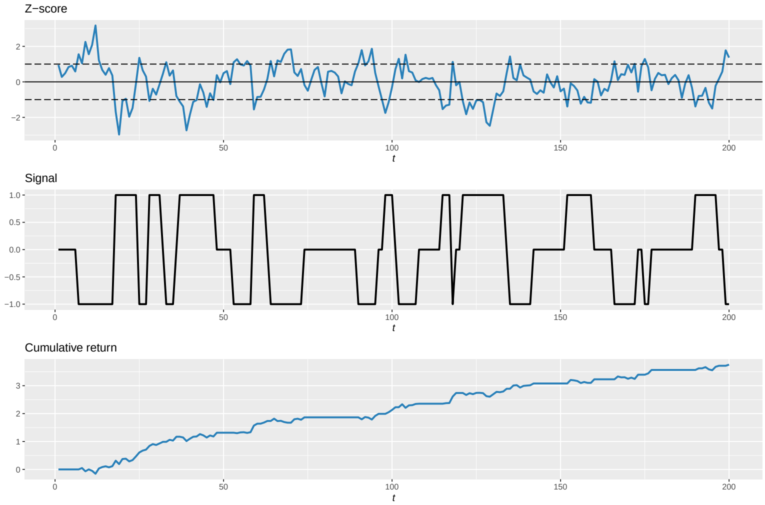


Illustration of pairs trading via the thresholded strategy on the spread:



Optimizing the Threshold

Thresholded Strategy:

- Buys when the z-score is below $-s_0$ and short-sells when it is above s_0.
- Unwinds the position after reverting to the equilibrium value of 0.
- In terms of the spread, the threshold is $s_0 \times \sigma$, where σ is the standard deviation of the spread.

Importance of Threshold Choice:

- Determines how often the position is closed (cashing a profit).
- Determines the size of the minimum profit.
- Total profit equals the number of trades times the profit of each trade.

Interpretation of Spread Difference:

- **Log-prices:** spread difference denotes the log-return of the profit.
- **Prices:** spread difference denotes the absolute profit (scaled with the initial budget).

Optimizing the Threshold

Total Profit Calculation:

- After N^{trades} successful trades, the total (uncompounded) profit is:

$$N^{\text{trades}} \times \sigma\, s_0$$

- Compounded profit could also be considered.

Optimizing the Threshold to maximize the total profit:

- **Parametric approach:** assumes a specific distribution for the spread and uses statistical methods to find the optimal threshold.
- **Non-parametric approach:** does not assume a specific distribution and uses empirical data to determine the optimal threshold.

Optimizing the Threshold: Parametric Approach

Parametric Model for Optimal Threshold:

- Assume the z-score follows a standard normal distribution: $z_t^{\text{score}} \sim \mathcal{N}(0, 1)$.
- Probability that z-score deviates from zero by s_0 or more: $1 - \Phi(s_0)$, where $\Phi(\cdot)$ is the cumulative distribution function (cdf) of the standard normal distribution.

Number of Tradable Events: For a time series path of T periods, the number of tradable events (in one direction) is approximated by:

$$T \times (1 - \Phi(s_0))$$

Total Profit:

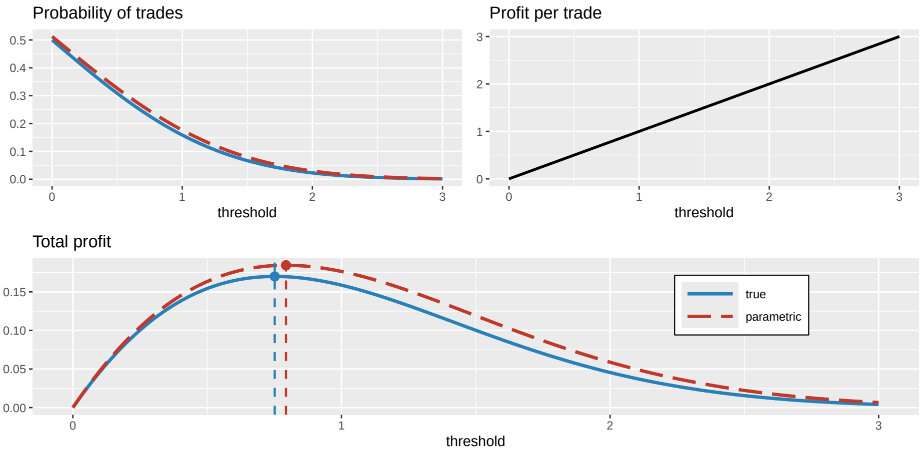
$$T\,(1 - \Phi(s_0)) \times \sigma\, s_0$$

Optimal Threshold: $s_0^\star$ to maximize the total profit:

$$s_0^\star = \arg\max_{s_0}\ (1 - \Phi(s_0)) \times s_0$$

Optimizing the Threshold: Parametric Approach

Calculation of optimum threshold in pairs trading via a parametric approach:



Optimizing the Threshold: Non-Parametric Approach

Data-Driven Approach:

- Does not rely on any model assumption.
- Uses available data to empirically count the number of tradable events.

Empirical Trading Frequency:

$$\bar{f}_j = \frac{1}{T} \sum_{t=1}^{T} 1\{z_t^{\text{score}} > s_{0j}\}$$

where

- z_t^{score} for $t = 1, \ldots, T$ are T observations of the z-score
- $s_{01}, \ldots, s_{0J}$ are J discretized threshold values
- $1\{\cdot\}$ is the indicator function.

Optimal Threshold:

$$s_0^\star = \underset{s_{0j} \in \{s_{01}, s_{02}, \ldots, s_{0J}\}}{\arg\max} \; s_{0j} \times \bar{f}_j$$

Optimizing the Threshold: Non-Parametric Approach

Noise Reduction:

- Empirical values $\bar{f}_j$ can be noisy.
- Reduce noise by leveraging the smoothness of the trading frequency function.
- Solve the least squares problem:

$$\underset{\boldsymbol{f}}{\text{minimize}} \quad \sum_{j=1}^{J}(f_j - \bar{f}_j)^2 + \lambda \sum_{j=1}^{J-1}(f_j - f_{j+1})^2$$

- First term measures the difference between noisy and smoothed values.
- Second term enforces smoothness, controlled by hyper-parameter λ.

Compact Notation:

$$\underset{\boldsymbol{f}}{\text{minimize}} \quad \|\boldsymbol{f} - \bar{\boldsymbol{f}}\|_2^2 + \lambda\|\boldsymbol{D}\boldsymbol{f}\|_2^2$$

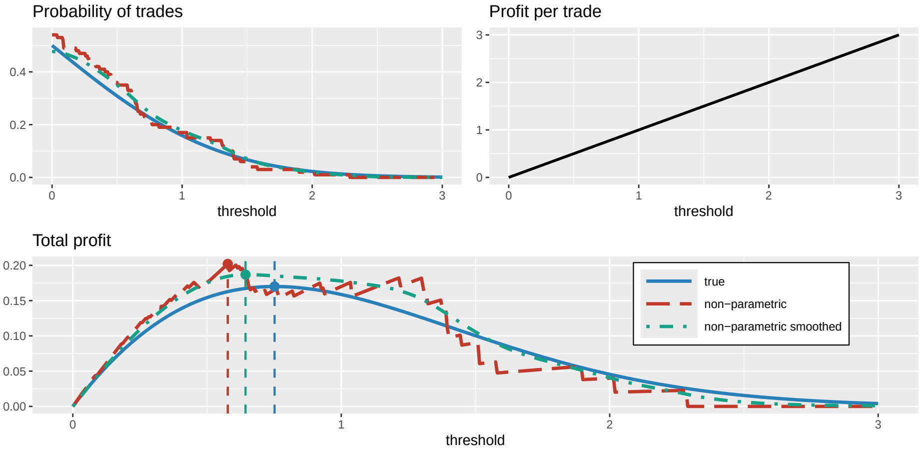
where $\boldsymbol{D}$ is the "difference matrix."

Closed-Form Solution:

$$\boldsymbol{f}^\star = \left(\boldsymbol{I} + \lambda\boldsymbol{D}^\mathsf{T}\boldsymbol{D}\right)^{-1}\bar{\boldsymbol{f}}$$

Optimizing the Threshold: Non-Parametric Approach

Calculation of optimum threshold in pairs trading via a non-parametric approach:



Numerical Experiments: Market Data EWA–EWC

EWA and EWC ETFs:

- Track the performance of the Australian and Canadian economies, respectively.
- Cointegration is present during most of the period, though occasionally lost as seen before.

***z*-score Calculation:**

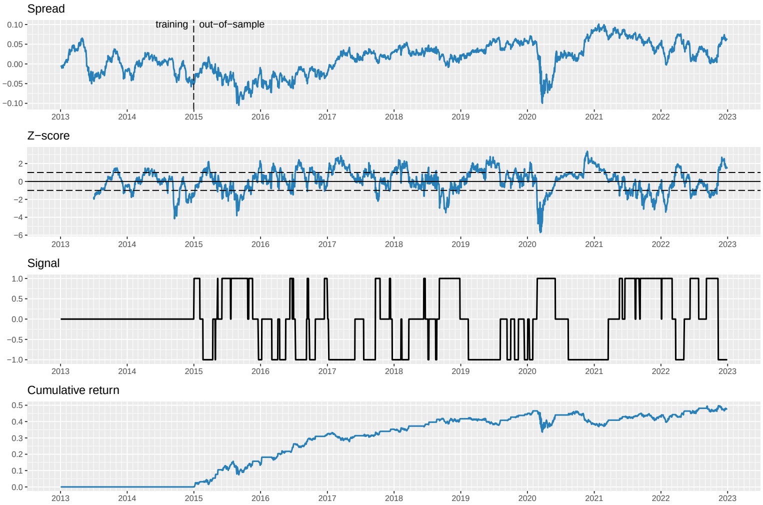
- Computed on a rolling window basis with a lookback period of 6 months.

Hedge Ratio:

- **Fixed hedge ratio:** Calculation of γ via least squares during the first 2 years of data.
  - Spread does not show a strong persistent cointegration relationship over the whole period.
  - *z*-score produces a more constant mean-reverting version due to rolling window adaptation.
- **Rolling hedge ratio:** Calculation of γ on a rolling-window basis with a lookback period of 2 years.
  - Improved spread compared to fixed least squares, showing more mean-reversion.
  - *z*-score further improves the mean-reverting version.
  - Cumulative return shows improvement due to rolling least squares approach.
  - Kalman filter provides even better results.

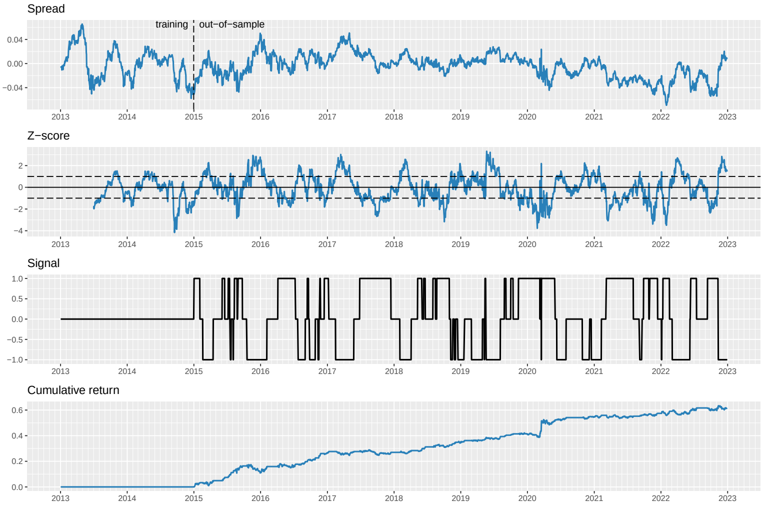
Numerical Experiments: Market Data EWA–EWC

Pairs trading on EWA–EWC with 6-month rolling z-score and 2-year **fixed least squares**:



Numerical Experiments: Market Data EWA–EWC

Pairs trading on EWA–EWC with 6-month rolling z-score and 2-year **rolling least squares**:



Numerical Experiments: Market Data KO–PEP

Pairs trading on Coca-Cola (KO) and Pepsi (PEP):

- Previous experiments indicate that KO and PEP do not seem to be cointegrated.

***z*-score Calculation:**

- Computed on a rolling window basis with lookback periods of 6 months and 1 month for faster adaptation.

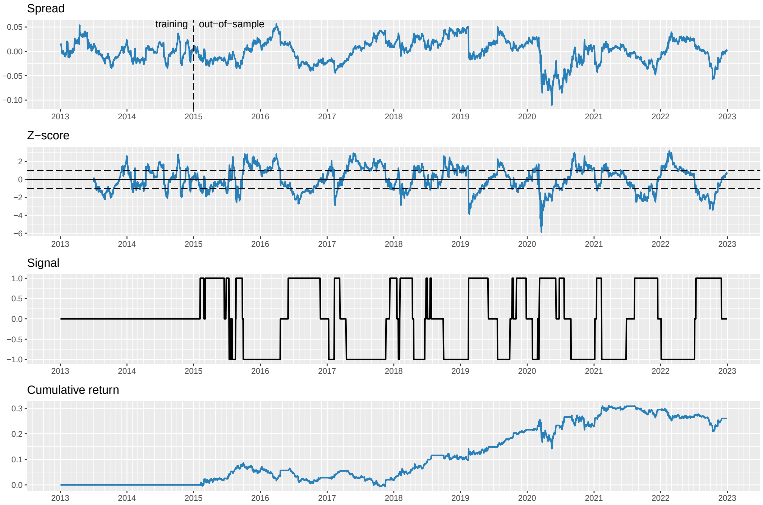
Hedge Ratio:

- **Rolling hedge ratio:** Calculation of γ on a rolling-window basis with a lookback period of 2 years.

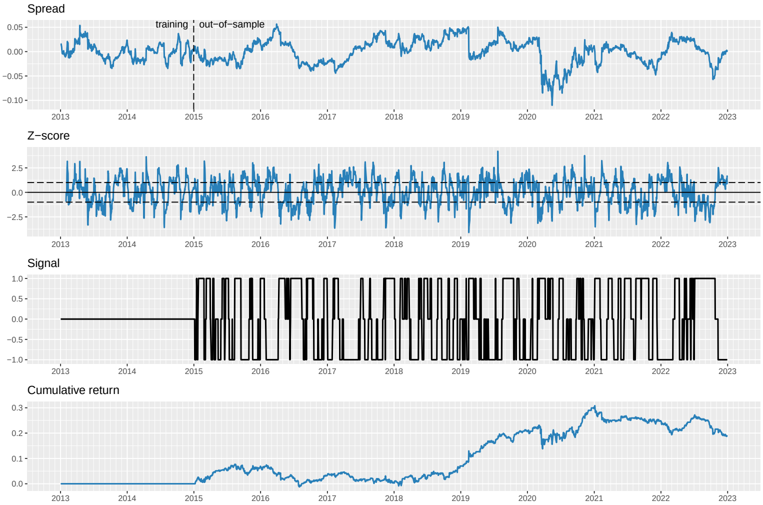
Observations:

- Improved spread compared to fixed least squares, showing more mean-reversion.
- Despite attempts to adapt the hedge ratio and z-score calculation, the lack of cointegration between KO and PEP results in poor trading performance.
- Faster adaptability of the z-score improves mean-reversion but does not translate into profitable trading.

Pairs trading on KO–PEP with **6-month rolling z-score** and 2-year rolling least squares:



Pairs trading on KO–PEP with **1-month rolling z-score** and 2-year rolling least squares:



Outline

Kalman for Pairs Trading

Mean-Reverting Spread Construction:

$$z_t = y_{1t} - \gamma\, y_{2t} = \mu + r_t$$

where

- γ: hedge ratio
- μ: mean of the spread
- r_t: zero-mean residual.

Estimation of Hedge Ratio and Mean:

- **Traditional method:**
  Employ least squares regression.
  Re-computed on a rolling window basis to adapt to changes over time.
- **Advanced method:**
  Use state space modeling and the Kalman filter for time-varying estimation (Feng and Palomar 2016).

Kalman for Pairs Trading

Kalman Filter:

- A powerful tool for estimating time-varying parameters.
- Provides a dynamic approach to update the hedge ratio γ and mean μ as new data becomes available.
- Helps in maintaining the mean-reverting property of the spread over time.

Advantages of Kalman Filter:

- Better adapts to changes in the relationship between the assets.
- Provides more accurate and timely estimates of the hedge ratio and mean.
- Enhances the performance of the pairs trading strategy by maintaining the mean-reverting nature of the spread.

Primer on Least Squares

- Dates back to 1795, used by Gauss to study planetary motions.
- Deals with the linear model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \epsilon$.
- Solves the problem:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$$

- Solution gives the least squares estimate:

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{A}^\mathsf{T}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^\mathsf{T}\boldsymbol{y}.$$

Spread Modeling via Least Squares

Spread Modeling Context:

- Fit $y_{1t} \approx \mu + \gamma\, y_{2t}$ based on T observations.
- LS formulation:

$$\underset{\mu,\gamma}{\text{minimize}} \quad \|\boldsymbol{y}_1 - (\mu\mathbf{1} + \gamma\, \boldsymbol{y}_2)\|_2^2$$

- Vectors $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ contain the T observations of the two time series.
- $\mathbf{1}$ is the all-one vector.

Practical Considerations:

- Hedge ratio γ and mean μ will change over time, denoted by γ_t and μ_t.
- LS solution must be re-computed on a rolling window basis (lookback window of past samples).

Kalman Filtering:

- Better handles the time-varying case.
- Provides a dynamic approach to update the hedge ratio γ_t and mean μ_t as new data becomes available.
- Enhances the performance of the pairs trading strategy by maintaining the mean-reverting nature of the spread over time.

Primer on Kalman

State Space Modeling:

- Provides a unified framework for time series analysis.
- Assumes system evolution is driven by unobserved or hidden values, measured indirectly through system output observations.
- Used for filtering, smoothing, and forecasting.

Kalman Filter:

- Efficient algorithm for state space models.
- Used by NASA in the Apollo program and now in various technological applications:
  Guidance, navigation, and control of vehicles (aircraft, spacecraft, maritime vessels).
  Time series analysis, signal processing, econometrics.
  Robotic motion planning and control, trajectory optimization.

Key references: (Anderson and Moore 1979), (Durbin and Koopman 2012), (Brockwell and Davis 2002), (Shumway and Stoffer 2017), (Harvey 1989), (Zivot, Wang, and Koopman 2004), (Tsay 2010), (Lütkepohl 2007), (Harvey and Koopman 2009).

Mathematical Formulation:

Linear Gaussian state space model with discrete-time $t = 1, \ldots, T$:

$$\boldsymbol{y}_t = \boldsymbol{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t \quad \text{(observation equation)}$$
$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{T}_t \boldsymbol{\alpha}_t + \boldsymbol{\eta}_t \quad \text{(state equation)}$$

- $\boldsymbol{y}_t$: observations over time with observation matrix $\boldsymbol{Z}_t$
- $\boldsymbol{\alpha}_t$: unobserved or hidden internal state with state transition matrix $\boldsymbol{T}_t$
- noise terms $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ are Gaussian distributed:
  $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{H})$
  $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q})$
- initial state: $\boldsymbol{\alpha}_1 \sim \mathcal{N}(\boldsymbol{a}_1, \boldsymbol{P}_1)$.

Primer on Kalman

Software Implementations:

- R packages: `KFAS` (Helske 2017), `MARSS` (Holmes, Ward, and Wills 2012).
- Python package: `filterpy`.

Parameter Estimation:

- Parameters $(\boldsymbol{Z}_t, \boldsymbol{T}_t, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{a}_1, \boldsymbol{P}_1)$ can be provided by the user or inferred from data using maximum likelihood methods.
- Efficient software implementations available for parameter fitting.

Kalman Filter Algorithm:

- Efficiently characterizes the distribution of the hidden state $\boldsymbol{\alpha}_t$ at time t.
- $\boldsymbol{\alpha}_{t|t-1}$: expected value given observations up to time $t-1$.
- $\boldsymbol{\alpha}_{t|t}$: expected value given observations up to time t.
- Computed using a "forward pass" algorithm from $t = 1$ to $t = T$ in a recursive manner, enabling real-time operation.

Spread Modeling via Kalman

State Space Modeling for Spread:

- Model $y_{1t} \approx \mu_t + \gamma_t \, y_{2t}$, where μ_t and γ_t change slowly over time.
- Hidden state: $\boldsymbol{\alpha}_t = (\mu_t, \gamma_t)$.
- State space model:

$$y_{1t} = \begin{bmatrix} 1 & y_{2t} \end{bmatrix} \begin{bmatrix} \mu_t \\ \gamma_t \end{bmatrix} + \epsilon_t$$

$$\begin{bmatrix} \mu_{t+1} \\ \gamma_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t \\ \gamma_t \end{bmatrix} + \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix},$$

- State transition matrix: $\boldsymbol{T} = \boldsymbol{I}$.
- Observation matrix: $\boldsymbol{Z}_t = \begin{bmatrix} 1 & y_{2t} \end{bmatrix}$.

Normalized Spread:

$$z_t = \frac{1}{1 + \gamma_{t|t-1}} \left(y_{1t} - \gamma_{t|t-1} \, y_{2t} - \mu_{t|t-1} \right),$$

where $\mu_{t|t-1}$ and $\gamma_{t|t-1}$ are the hidden states estimated by the Kalman filter.

Numerical Experiments: Market Data EWA–EWC

Pairs Trading Experiments (2013-2022):

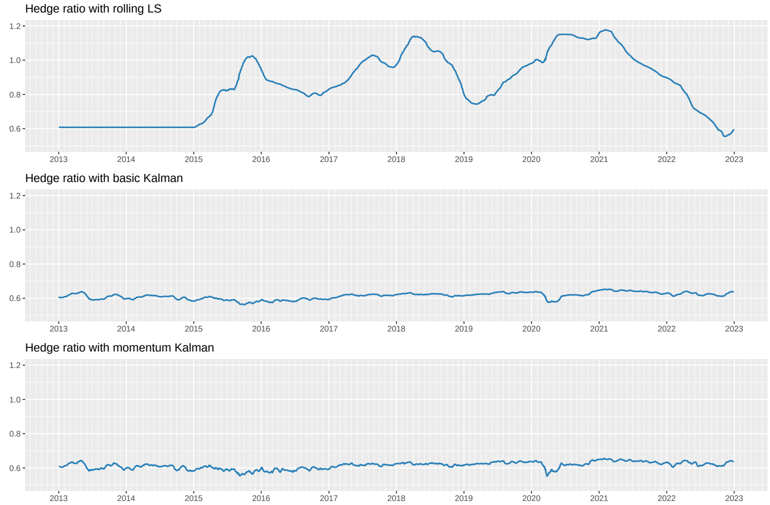
- z-score computed on a rolling window basis with a lookback period of 6 months.
- Pairs trading implemented via the thresholded strategy with a threshold of $s_0 = 1$.
- Three methods to track the hedge ratio over time:
  1. Rolling least squares with a lookback period of two years.
  2. Basic Kalman filter.
  3. Kalman filter with momentum in the modeling.
- Parameters fixed but could be optimized; state space model parameters can be learned via maximum likelihood estimation.

Observations:

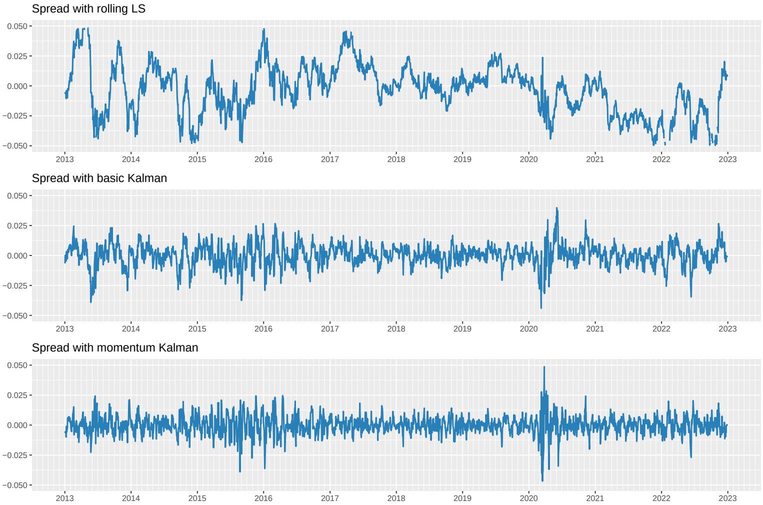
- Kalman-based methods provide more stable hedge ratios and better mean-reverting spreads.
- Result in higher cumulative returns compared to rolling least squares:
  - Rolling least squares: final value 0.6.
  - Basic Kalman filter: final value 2.0.
  - Kalman filter with momentum: final value 3.2.

Numerical Experiments: Market Data EWA–EWC

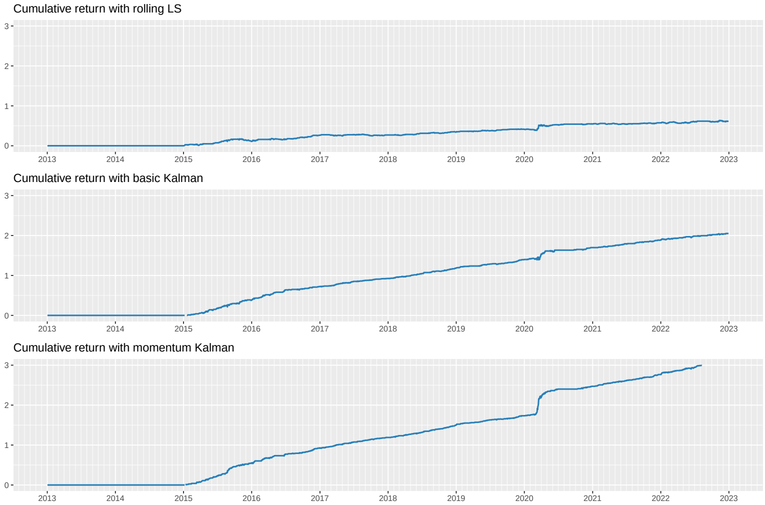
Tracking of hedge ratio for pairs trading on EWA–EWC:



Spread for pairs trading on EWA–EWC:



Cumulative return for pairs trading on EWA–EWC:



Numerical Experiments: Market Data KO–PEP

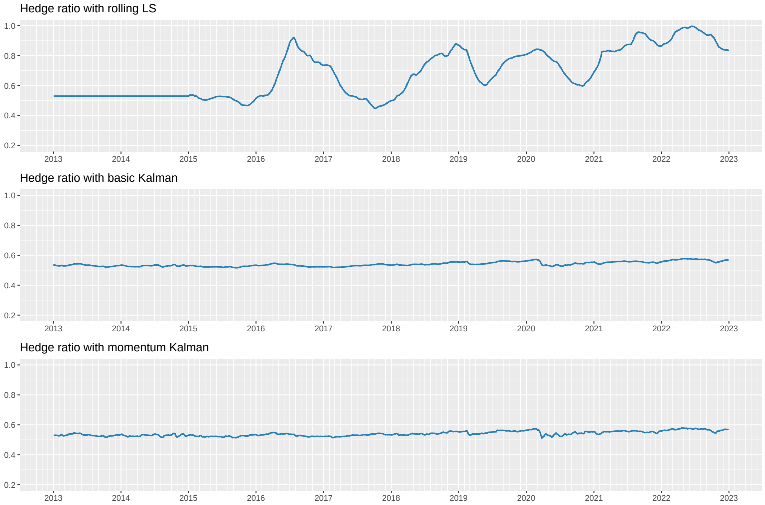
Pairs Trading with Coca-Cola (KO) and Pepsi (PEP):

- Previous cointegration tests indicate they are not cointegrated.
- Previous trading experiments showed dubious profitability.
- Now we use Kalman-based methods to see if the situation improves.

Observations:

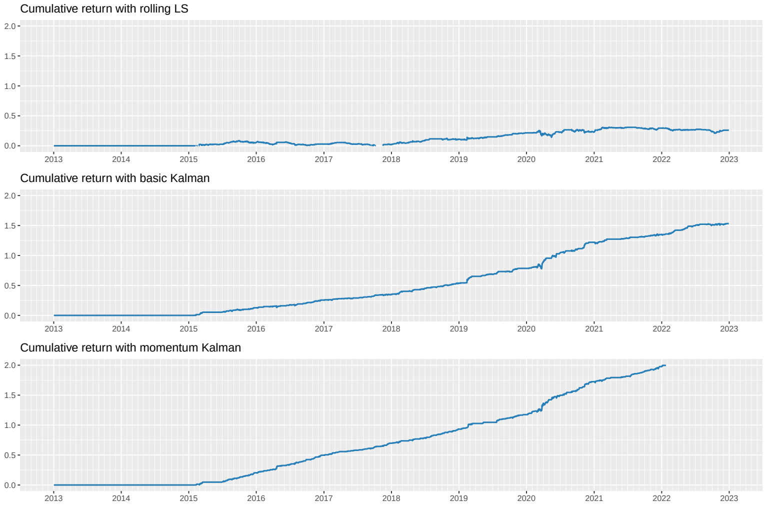
- Kalman-based methods provide more stable hedge ratios and better mean-reverting spreads, especially during the big change in early 2020 (likely due to COVID-19 pandemic).
- Significant differences in spreads among the three methods. Early 2020:
  - Rolling least squares: loses tracking, cointegration clearly lost.
  - Basic Kalman: tracks after a momentary loss, reflected in a shock on the spread.
  - Kalman with momentum: tracks much better.
- Kalman-based methods: better performance and controlled drawdown.

Tracking of hedge ratio for pairs trading on KO–PEP:



Spread for pairs trading on KO–PEP:

Cumulative return for pairs trading on KO–PEP:



Outline

Statistical Arbitrage (StatArb)

Pairs Trading:

- Focuses on discovering cointegration and tracking the cointegration relationship between pairs of assets.
- Can be extended to more than two assets for more flexibility.
- This extension is generally referred to as *statistical arbitrage* or *StatArb*.

Cointegration for Multiple Assets:

- Follows the same idea as for pairs: construct a linear combination of multiple time series such that the combination is mean-reverting.
- Mathematical modeling becomes more involved to capture multiple cointegration relationships.

Least Squares

Least Squares for Cointegration:

- Can still be used to determine the cointegration relationship for $K > 2$ time series.
- Requires choosing one time series to be regressed by the others.

Least Squares Formulation:

- Suppose we want to fit $y_{1t} \approx \mu + \sum_{k=2}^{K} \gamma_k \, y_{kt}$ based on T observations.
- The least squares formulation is:

$$\underset{\mu,\gamma}{\text{minimize}} \quad \left\| \boldsymbol{y}_1 - (\mu \mathbf{1} + \boldsymbol{Y}_2 \gamma) \right\|_2^2 ,$$

where

- $\boldsymbol{y}_1$: vector containing T observations of the first time series
- $\boldsymbol{Y}_2$: matrix containing T observations of the remaining $K-1$ time series columnwise
- $\gamma \in \mathbb{R}^{K-1}$: vector containing the $K-1$ hedge ratios.

Normalized Portfolio:

$$\boldsymbol{w} = \frac{1}{1 + \|\gamma\|_1} \begin{bmatrix} 1 \\ -\gamma \end{bmatrix}$$

which leads to the normalized spread:

$$z_t = \boldsymbol{w}^\mathsf{T} \boldsymbol{y}_t$$

Limitations:

- This approach produces a single cointegration relationship (other cointegration relationships may go unnoticed).
- Requires choosing one time series (out of the K possible ones) to be regressed.
- One approach is to iteratively capture more cointegration relationships orthogonal to the previously discovered ones.

Sophisticated VECM Modeling:

- Discovery of multiple cointegration relationships is better achieved by the more sophisticated VECM (Vector Error Correction Model) modeling described next.

VECM

Multivariate Time Series Models:

- Commonly used for log-prices of N assets.
- Based on first-order difference: $\Delta \boldsymbol{y}_t = \boldsymbol{y}_t - \boldsymbol{y}_{t-1}$.
- Uses a vector autoregressive (VAR) model of order p:

$$\Delta \boldsymbol{y}_t = \phi_0 + \sum_{i=1}^{p} \boldsymbol{\Phi}_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{\epsilon}_t,$$

- Differencing makes the model stationary but may destroy some structure in the data.

Vector Error Correction Model (VECM):

- Proposed by Engle and Granger (1987) to apply the VAR model without differencing.
- Potential danger of lack of stationarity.
- VECM model:

$$\Delta \boldsymbol{y}_t = \phi_0 + \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1} \tilde{\boldsymbol{\Phi}}_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{\epsilon}_t,$$

- The term $\boldsymbol{\Pi} \boldsymbol{y}_{t-1}$ must be stationary since $\Delta \boldsymbol{y}_t$ is stationary.

VECM

Matrix $\mathbf{\Pi}$:

- Crucial for guaranteeing stationarity of $\mathbf{\Pi}\boldsymbol{y}_t$.
- Generally of low rank, decomposable as:

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}^{\mathsf{T}}$$

- $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^{N \times K}$ with K columns (rank of $\mathbf{\Pi}$).
- Nonstationary series $\boldsymbol{y}_t$ becomes stationary after multiplication with $\boldsymbol{\beta}^{\mathsf{T}}$.
- Each column of $\boldsymbol{\beta}$ produces a different cointegration relationship.

Cases Based on the Rank of $\mathbf{\Pi}$:

- $K = N$: $\boldsymbol{y}_t$ is already stationary (rare in practice).
- $K = 0$: $\boldsymbol{y}_t$ is not cointegrated (VECM reduces to a VAR model).
- $1 < K < N$: provides K different cointegration relationships.

Johansen's Test:

- Tests the value of the rank of the matrix $\mathbf{\Pi}$ in VECM time series modeling.

Optimum Mean-Reverting Portfolio

Cointegration Relationships for Pairs Trading and Statistical Arbitrage:

- Least squares and VECM modeling can be used to obtain cointegration relationships.
- VECM provides K different cointegration relationships in the columns of matrix $\boldsymbol{\beta} \in \mathbb{R}^{N \times K}$.
- K different pairs trading strategies can be run in parallel, exploiting all K directions in the N-dimensional space.

Optimization-Based Approach:

- Design the portfolio to maximize the zero-crossing rate and the variance of the spread.
- Proxies for zero-crossing rate produce various problem formulations (d'Aspremont 2011; Cuturi and d'Aspremont 2013, 2016).

Combined Approach:

- Use VECM modeling to define a *cointegration subspace*.
- Optimize portfolios within this subspace for better spreads (Zhao and Palomar 2018; Zhao, Zhou, and Palomar 2019).

Whole Procedure for Combined Approach:

1. **Cointegration relationships:** from K columns of matrix β:
$$\beta_k \in \mathbb{R}^N, \qquad k = 1, \dots, K.$$
2. **Normalized portfolios:** construct K portfolios:
$$\mathbf{w}_k = \frac{1}{\|\beta_k\|_1}\beta_k, \qquad k = 1, \dots, K.$$
3. **Compute spreads:** from original time series $\mathbf{y}_t \in \mathbb{R}^N$:
$$z_{kt} = \mathbf{w}_k^T \mathbf{y}_t, \qquad k = 1, \dots, K$$
or, more compactly:
$$\mathbf{z}_t = [\mathbf{w}_1 \dots \mathbf{w}_K]^T \, \mathbf{y}_t \in \mathbb{R}^K.$$
4. **Optimize spread portfolio:** optimize a K-dimensional portfolio $\mathbf{w}_z \in \mathbb{R}^K$ on the spreads $\mathbf{z}_t$:
$$\mathbf{w}^{\text{overall}} = [\mathbf{w}_1 \dots \mathbf{w}_K] \times \mathbf{w}_z.$$

Optimum Mean-Reverting Portfolio

Optimization of the Spread Portfolio $\boldsymbol{w}_z$:

- Goal: optimize some proxy of the zero-crossing rate while controlling spread variance.
- Define lagged covariance matrices of the spreads:

$$\boldsymbol{M}_i = \mathbb{E}\left[(\boldsymbol{z}_t - \mathbb{E}[\boldsymbol{z}_t])\,(\boldsymbol{z}_{t+i} - \mathbb{E}[\boldsymbol{z}_{t+i}])^{\mathsf{T}}\right], \qquad i = 0, 1, 2, \ldots$$

- Variance of the resulting spread: $\boldsymbol{w}_z^{\mathsf{T}}\boldsymbol{M}_0\boldsymbol{w}_z$.

Proxies for Zero-Crossing Rate:

- **Predictability statistic:**

$$\mathrm{pre}(\boldsymbol{w}_z) = \frac{\boldsymbol{w}_z^{\mathsf{T}}\boldsymbol{M}_1^{\mathsf{T}}\boldsymbol{M}_0^{-1}\boldsymbol{M}_1\boldsymbol{w}_z}{\boldsymbol{w}_z^{\mathsf{T}}\boldsymbol{M}_0\boldsymbol{w}_z}.$$

- **Portmanteau statistic:**

$$\mathrm{por}(\boldsymbol{w}_z) = \sum_{i=1}^{p}\left(\frac{\boldsymbol{w}_z^{\mathsf{T}}\boldsymbol{M}_i\boldsymbol{w}_z}{\boldsymbol{w}_z^{\mathsf{T}}\boldsymbol{M}_0\boldsymbol{w}_z}\right)^2.$$

- **Crossing statistic:**

$$\mathrm{cro}(\boldsymbol{w}_z) = \frac{\boldsymbol{w}_z^{\mathsf{T}}\boldsymbol{M}_1\boldsymbol{w}_z}{\boldsymbol{w}_z^{\mathsf{T}}\boldsymbol{M}_0\boldsymbol{w}_z}.$$

Optimum Mean-Reverting Portfolio

Mean-Reverting Portfolio Optimization:

Optimize zero-crossing proxy while fixing spread variance:

$$
\begin{array}{ll}
\underset{\boldsymbol{w}_z}{\text{minimize}} & \boldsymbol{w}_z^\mathsf{T}\boldsymbol{M}_1\boldsymbol{w}_z + \eta\sum_{i=2}^{p}\left(\boldsymbol{w}_z^\mathsf{T}\boldsymbol{M}_i\boldsymbol{w}_z\right)^2\\
\text{subject to} & \boldsymbol{w}_z^\mathsf{T}\boldsymbol{M}_0\boldsymbol{w}_z \geq \nu\\
& \boldsymbol{w}_z \in \mathcal{W},
\end{array}
$$

where ν, η are hyper-parameters, and $\mathcal{W}$ denotes portfolio constraints, such as:

- $\|\boldsymbol{w}_z\|_2 = 1$ to avoid numerical issues (Cuturi and d'Aspremont 2013).
- Sparsity constraint $\|\boldsymbol{w}_z\|_0 = k$ (Cuturi and d'Aspremont 2016).
- Budget/market exposure constraint $\mathbf{1}^\mathsf{T}\boldsymbol{w}_z = 1/0$ (Zhao and Palomar 2018).
- Leverage constraint $\|\boldsymbol{w}_z\|_1 = 1$ (Zhao, Zhou, and Palomar 2019).
- Leverage constraint on the overall portfolio:

$$\|\left[\boldsymbol{w}_1 \ldots \boldsymbol{w}_K\right] \times \boldsymbol{w}_z\|_1 = 1.$$

Numerical Experiments: Market Data SPY–IVV–VOO

Illustration of Multiple Cointegration Relationships via VECM Modeling:

- Consider three ETFs tracking the S&P 500 index: SPY, IVV, and VOO.
- Johansen's test indicates two cointegration relationships, exploitable via statistical arbitrage.

Results:

Next figure and table shows results for:

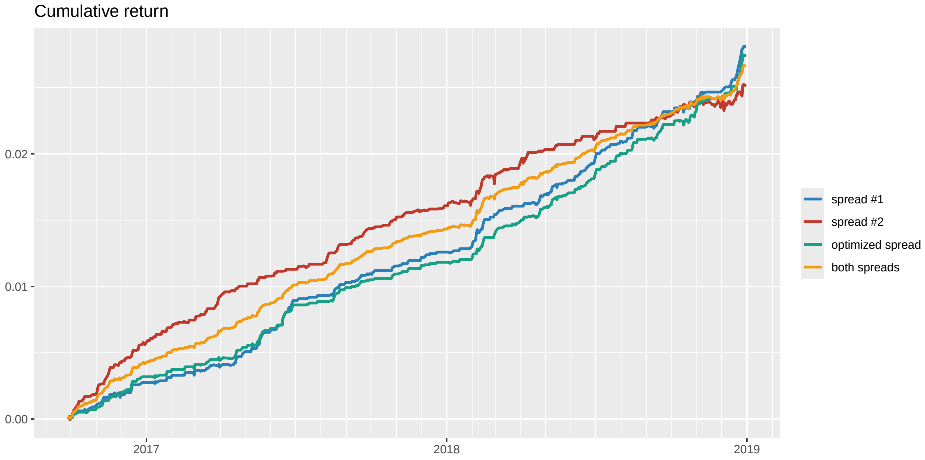
1. First (strongest) spread.
2. Second (weaker) spread.
3. Optimized spread within the cointegrated subspace.
4. Both spreads in parallel (allocating half of the budget to each spread).

Observations:

- Strongest spread performs better than the second spread.
- Optimized spread does not offer an improvement in this case (may offer benefits for larger dimensionality of the cointegrated subspace).
- Using both spreads in parallel provides a more steady cumulative return (better Sharpe ratio) due to diversity gain.

Numerical Experiments: Market Data SPY–IVV–VOO

Cumulative return for pairs trading on SPY–IVV–VOO: single spreads, both in parallel, and optimized spread:



Numerical Experiments: Market Data SPY–IVV–VOO

Sharpe ratios for pairs trading on SPY–IVV–VOO: single spreads, both in parallel, and optimized spread:

Spread	Sharpe ratio
spread #1	6.78
spread #2	5.39
optimized spread	6.75
both spreads	8.37

Outline

Summary

Pairs trading and statistical arbitrage are market-neutral strategies exploiting relative asset values. Key concepts include:

- **Mean-reversion**: Time series fluctuate around a long-term average, enabling a buy-low, sell-high strategy.
- **Cointegration**: Assets that are not mean-reverting individually but become so when combined.
- **Pairs trading**: Uses two cointegrated assets to create a market-neutral, mean-reverting synthetic asset, contrasting with trend-following momentum strategies.
- **Implementation**: Involves discovering cointegrated assets (statistical tests), tracking relationships (rolling least squares or Kalman filtering), and executing trades (threshold strategy).
- **Kalman filtering**: Tracks cointegration over time using state space models.
- **Statistical arbitrage**: Extends pairs trading to multiple assets, requiring multivariate modeling (VECM) to identify cointegration.

Anderson, B. D. O., and J. B. Moore. 1979. *Optimal Filtering*. Englewood Cliffs: Prentice-Hall.

Avellaneda, M., and J.-H. Lee. 2010. "Statistical Arbitrage in the US Equities Market." *Quantitative Finance* 10 (7): 761–82.

Brockwell, P. J., and R. A. Davis. 2002. *Introduction to Time Series and Forecasting*. Springer.

Chan, E. P. 2008. *Quantitative Trading: How to Build Your Own Algorithmic Trading Business*. Wiley.

———. 2013. *Algorithmic Trading: Winning Strategies and Their Rationale*. Wiley.

Clegg, M. 2014. "On the Persistence of Cointegration in Pairs Trading." *SSRN Electronic Journal*. https://dx.doi.org/10.2139/ssrn.2491201.

References II

Clegg, M., and C. Krauss. 2018. "Pairs Trading with Partial Cointegration." *Quantitative Finance* 18 (1): 121–38.

Cuturi, M., and A. d'Aspremont. 2013. "Mean Reversion with a Variance Threshold." In *Proceedings of the International Conference on Machine Learning (ICML)*, 28:271–79.

———. 2016. "Mean-Reverting Portfolios: Tradeoffs Between Sparsity and Volatility." In *Financial Signal Processing and Machine Learning*, edited by A. N. Akansu, S. R. Kulkarni, and D. M. Malioutov, 23–40. Wiley.

d'Aspremont, A. 2011. "Identifying Small Mean-Reverting Portfolios." *Quantitative Finance* 11 (3): 351–64.

Durbin, J., and S. J. Koopman. 2012. *Time Series Analysis by State Space Methods*. 2nd ed. Oxford University Press.

References III

Ehrman, D. S. 2006. *The Handbook of Pairs Trading: Strategies Using Equities, Options, and Futures*. John Wiley & Sons.

Elliott, R. J., J. Van Der Hoek, and W. P. Malcolm. 2005. "Pairs Trading." *Quantitative Finance* 5 (3): 271–76.

Engle, R. F., and C. W. J. Granger. 1987. "Co-Integration and Error Correction: Representation, Estimation, and Testing." *Econometrica: Journal of the Econometric Society*, 251–76.

Feng, Y., and D. P. Palomar. 2016. *A Signal Processing Perspective on Financial Engineering*. Foundations and Trends in Signal Processing, Now Publishers.

Gatev, E., W. N. Goetzmann, and K. G. Rouwenhorst. 2006. "Pairs Trading: Performance of a Relative-Value Arbitrage Rule." *Review of Financial Studies* 19 (3): 797–827.

References IV

Harvey, A. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Harvey, A., and S. J. Koopman. 2009. "Unobserved Components Models in Economics and Finance: The Role of the Kalman Filter in Time Series Econometrics." *IEEE Control Systems Magazine* 29 (6): 71–81.

Helske, J. 2017. "KFAS: Exponential Family State Space Models in R." *Journal of Statistical Software* 78 (10): 1–39.

Holmes, E. E., E. J. Ward, and K. Wills. 2012. "MARSS: Multivariate Autoregressive State-Space Models for Analyzing Time-Series Data." *The R Journal* 4 (1): 11–19.

Johansen, S. 1991. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica: Journal of the Econometric Society*, 1551–80.

———. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

Lütkepohl, H. 2007. *New Introduction to Multiple Time Series Analysis*. Springer.

Palomar, D. P. 2025. *Portfolio Optimization: Theory and Application*. Cambridge University Press.

Shumway, R. H., and D. S. Stoffer. 2017. *Time Series Analysis and Its Applications*. 4th ed. Springer.

Tsay, R. S. 2010. *Analysis of Financial Time Series*. 3rd ed. John Wiley & Sons.

Vidyamurthy, G. 2004. *Pairs Trading: Quantitative Methods and Analysis*. John Wiley & Sons.

Zhao, Ziping, and Daniel P. Palomar. 2018. "Mean-Reverting Portfolio with Budget Constraint." *IEEE Transactions on Signal Processing* 66 (9): 2342–57.

References VI

Zhao, Ziping, Rui Zhou, and Daniel P. Palomar. 2019. "Optimal Mean-Reverting Portfolio with Leverage Constraint for Statistical Arbitrage in Finance." *IEEE Transactions on Signal Processing* 67 (7): 1681–95.

Zivot, E., J. Wang, and S. J. Koopman. 2004. "State Space Modeling in Macroeconomics and Finance Using SsfPack for S+FinMetrics." In *State Space and Unobserved Component Models: Theory and Applications*, edited by A. Harvey, S. J. Koopman, and N. Shephard, 284–335. Cambridge University Press.