Adaptive Pairs Trading using Econometric Models and Bayesian Optimization

Group 11
QF603

Prepared by:
Htoo Myat Naing, Malcolm Yong Junxi, Tan Wei Lin Gregory, Karan Rajesh Chavan, Aryan Goel

Table of Contents

Project Overview

Overview of Pairs Trading

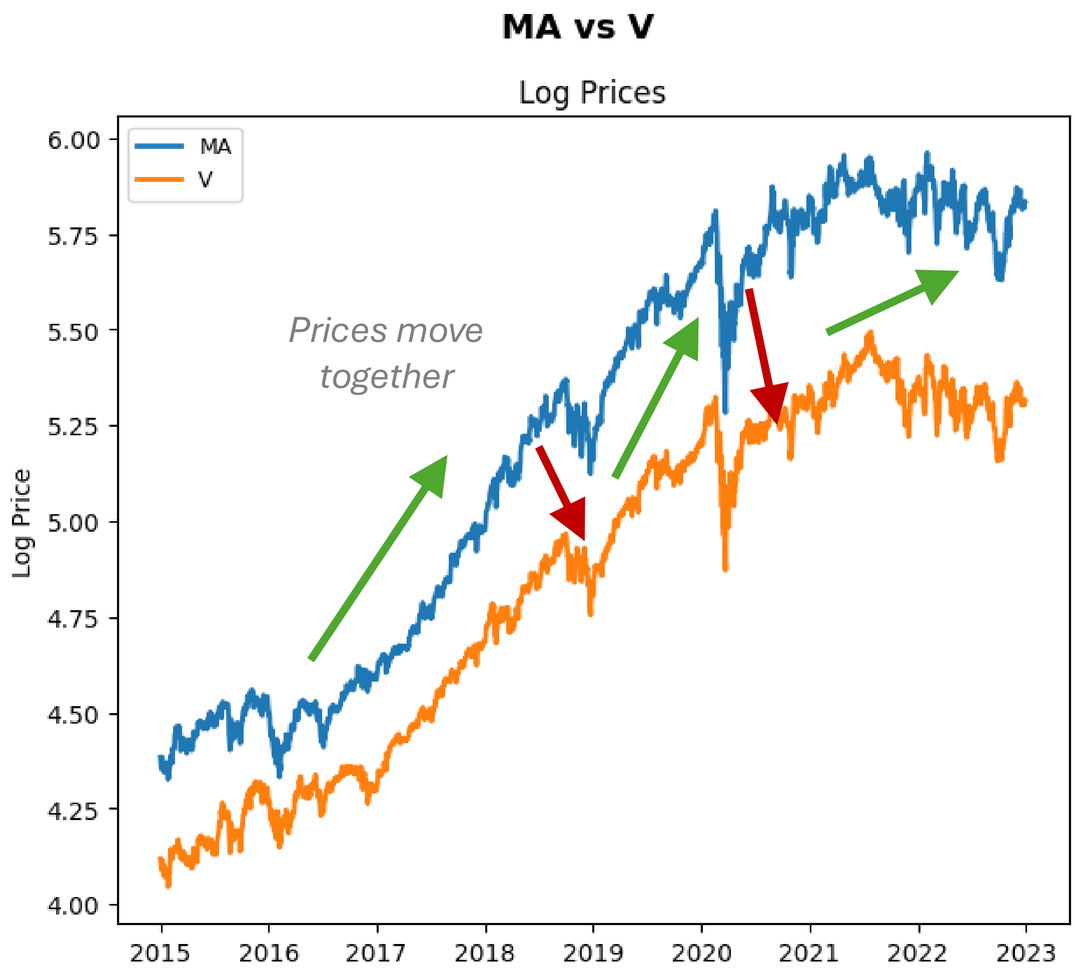
What is Pairs Trading?

- A **market-neutral strategy** that exploits **relative mispricing between two assets** with a **historically stable long-run equilibrium relationship (cointegrated)**

- When the price difference (spread) between the two cointegrated assets widen, go **long on undervalued asset** and **short-sell the overvalued asset**

- As price **reverts toward equilibrium**, both positions converge, yielding profit independent of market direction

"Buy the cheap,
Sell the expensive"



Project Overview

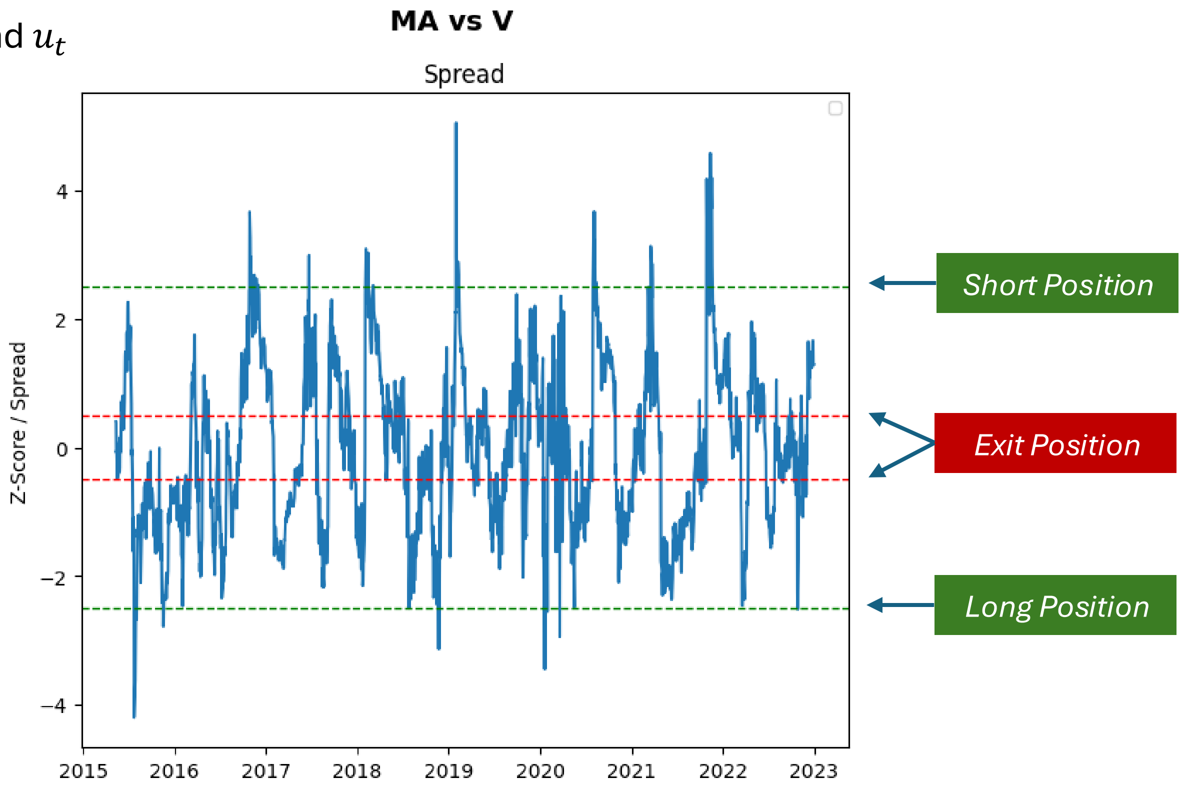
Implementation of Pairs Trading

A Basic Pair Trading Strategy

- Compare log-prices of two assets, x_1 and x_2 at time t to determine spread u_t

$$u_t = x_{1t} - \beta \cdot x_{2t}$$

- u_t is expected to be mean-reverting with mean μ
- β represents the normalized dollar-value of x_1 to hedge against x_2



Example Trading Strategy

- Set a threshold θ, to enter long/short when spread deviates from μ
- **Long signal:** spread is low, $u_t < \mu - \theta$
- **Short signal:** spread is high, $u_t > \mu + \theta$
- Exit position: spread converges back to μ

Risks

- Assumption that spread will remain mean-reverting may not hold due to switching of market regimes, fundamental divergences, etc.

Project Workflow

A Detailed Sequence to the Pairs Trading Approach

Screening for Cointegrated Pairs

*Using tickers from S&P 500

ADF and **KPSS** to determine order of integration

Engle-Granger, **Phillips-Ouliaris** and **Johansen** tests for pair cointegration

Model Fitting

Fit **ECM**, **VECM** and **Kalman Filter** to capture price relationships

Signal Generation and Backtesting

Using validation data, trade spread using predefined entry and exit criteria, inclusive of costs and margin constraints

Strategy Optimization

Use **Bayesian Optimization** to select model parameters based on maximizing Sharpe ratio

Final Testing

Apply optimized parameters on test data (out-of-sample) to test performance of models

Data & Preprocessing

Data Source, Data Cleaning and Feature Preparation

Data Universe

- **Universe**: S&P 500 (as of Oct 2025)
- **Date Range**: Jan 2015 to Oct 2025
- **Frequency**: Daily
- **Source**: Yahoo Finance (yfinance)
- **Ticker-level metadata**:
  - Industry
  - Sector
  - Floored $\log_{10}$(Market Cap)
  - Floored $\log_{1000}$(Dollar Volume)

	Symbol	yf_ticker	industry	sector	log10_market_cap	log1000_volume
0	A	A	Diagnostics & Research	Healthcare	10	2
1	AAPL	AAPL	Consumer Electronics	Technology	12	3
2	ABBV	ABBV	Drug Manufacturers - General	Healthcare	11	3
4	ABT	ABT	Medical Devices	Healthcare	11	2
5	ACGL	ACGL	Insurance - Diversified	Financial Services	10	2
...	...	...	...	...	...	...
497	XYL	XYL	Specialty Industrial Machinery	Industrials	10	2
499	YUM	YUM	Restaurants	Consumer Cyclical	10	2
500	ZBH	ZBH	Medical Devices	Healthcare	10	2
501	ZBRA	ZBRA	Communication Equipment	Technology	10	2
502	ZTS	ZTS	Drug Manufacturers - Specialty & Generic	Healthcare	10	2

Data Cleaning and Feature Preparation

- Used **Adjusted Close** prices to remove impact of dividends and stock splits
- Filtered for stocks with **no missing prices** and **no zero trading volume** across entire date range, ensuring active and continuous trading history
- Transformed prices into **log prices** to stabilize variance and linearize price shifts for modelling
- Split data into **Train**, **Validation** and **Test** sets for model training, tuning and evaluation purposes

Train	Jan 2015 – Dec 2019
Validation	Jan 2020 – Dec 2022
Test	Jan 2023 – Oct 2025

Screening for Cointegrated Pairs

Finding order of integration for each ticker and creating ticker pairs for cointegration testing

SMU SINGAPORE MANAGEMENT UNIVERSITY

Finding Order of Integration

- Iteratively run **ADF** and **KPSS test** at **5% level of significance** on log prices differenced to the i^{th} order, starting from $i = 0$, until H_0 is rejected (i.e., the i^{th}-differenced time series is stationary)
- **Retain tickers that are I(1)** – non-stationary in levels but stationary after first differencing
    - Ensures valid cointegration testing, as mixed integration orders would lead to spurious relationships

	Symbol	industry	sector	log10_market_cap	log1000_volume	log_integration
0	A	Diagnostics & Research	Healthcare	10	2	1
1	AAPL	Consumer Electronics	Technology	12	3	1
2	ABBV	Drug Manufacturers - General	Healthcare	11	3	1
4	ABT	Medical Devices	Healthcare	11	2	1
5	ACGL	Insurance - Diversified	Financial Services	10	2	1
...	...	...	...	...	...	...
497	XYL	Specialty Industrial Machinery	Industrials	10	2	1
499	YUM	Restaurants	Consumer Cyclical	10	2	1
500	ZBH	Medical Devices	Healthcare	10	2	1
501	ZBRA	Communication Equipment	Technology	10	2	1
502	ZTS	Drug Manufacturers - Specialty & Generic	Healthcare	10	2	1

**Retain only I(1)*

Grouping Tickers by Economic Similarity

- Grouped filtered tickers using metadata such as industry, sector, floored $\log_{10}$(Market Cap) and floored $\log_{1000}$(Volume)
- **Removed pairs that did not share a common industry or sector** to focus on economically related companies (more likely to exhibit genuine long-run relationships)

	ticker1	ticker2	sector	industry	log10_market_cap	log1000_volume
30	ADBE	ADP	Technology	Software - Application	11	NaN
31	ADI	ADP	Technology	NaN	11	2
32	ADM	ADP	NaN	NaN	NaN	2

**Row 32 will be removed*

Screening for Cointegrated Pairs

Identification of potential cointegrated pairs using multiple hypothesis tests

Tests	Key Idea	Purpose	Hypothesis Testing	Limitation / Adjustment
Engle-Granger (EG)	Estimate one series Y_t using an OLS on the other X_t. Run an ADF test on residuals $\hat{u}_t$ and compare test stat to EG critical values.	Conceptually intuitive and widely used for bivariate cointegration detection	H_0: Residuals have serial correlation; non-cointegrated H_1: Residuals are stationary, both series are cointegrated	**Directional Bias:** Results differ based on the specified exogenous variable. Hence, we run the EG test in both directions.
Phillips-Ouliaris (PO) Pz Test	Check for evidence of divergence in the time T-scaled ratio between the long-run covariance and the short-term variance $T \cdot tr(\widehat{\Omega} M_{zz})$	System-based test treats all variables as endogenous, removing directional biases.	H_0: Variables are random walks; non-cointegrated H_1: At least one stationary linear combination; cointegrated	**Specificity:** Unable to determine specific cointegrating rank, r^* for ≥2 assets.
Johansen Trace Test (JOT)	Identify the rank, r^* of the VECM-estimated long-run impact matrix Π, which represents the number of cointegrating relationships.	Same benefits as PO test, and can state the exact number of cointegrating vectors, which is useful in extending beyond pairs.	For $r = 0$ until we do not reject H_0, $H_0: r^* \leq r$ $H_1: r^* > r$ $r^* = r$ for the first r we do not reject	**Sensitivity to Parameters:** results can be sensitive to specified lag parameter. We use a VAR model to select based on BIC.

Note: Standard stationarity tests (standard ADF critical values) cannot be applied to regressed residuals because they are estimated (e.g., by OLS), not observed.
For all model estimations, we include a constant term to remove bias from the determined long-run equilibrium.

Screening for Cointegrated Pairs

Identification of potential cointegrated pairs using multiple hypothesis tests

Selection Methodology

- To ensure robustness, a pair is considered cointegrated if it passes all tests at the 95% level of significance
- This reduces false positives and confirms that long-run relationships holds under multiple testing perspectives
- **In short**: *only pairs consistently mean-reverting across all frameworks are traded*

Choosing A Pair

- From our screening methodology, we identified the following pair **PODD/RMD**, which passed all hypothesis tests.
- Insulet Corp (PODD) is a medical device company engaging in development, manufacture and sale of insulin products.
- ResMed Inc. (RMD) through its subsidiaries, engages in development, manufacturing, and distribution of medical devices and apps that treat respiratory disorders and other diseases.

Test	Test Stat	95% CV	99% CV	P-Value	Results
EG ($y =$ PODD)	-5.5821	-3.3424	-3.9068	1.2e-5	Cointegrated
EG ($y =$ RMD)	-5.5646	-3.3424	-3.9068	8.0e-6	Cointegrated
PO	98.1368	53.9652	69.9581	7.6e-4	Cointegrated
JOT ($r = 0$)	32.3525	15.4943	19.9349	-	Cointegrated, $r = 1$
JOT ($r = 1$)	0.4260	3.8415	6.6349	-	

PODD/RMD is cointegrated

Cointegration Modelling

Error Correction Model

Error Correction Model (ECM)

$$\Delta Y_t = \gamma + \beta_1 \Delta X_t + \alpha(Y_{t-1} - \beta_0 - \beta_1 X_{t-1}) + \epsilon_t$$

where

$\mathbf{Y_t, X_t}$: log prices assets

$\mathbf{\Delta Y_t, \Delta X_t}$: log returns

$\boldsymbol{\alpha}$: speed of adjustment

$\boldsymbol{\gamma}$: short-run constant

$\mathbf{Y_{t-1} - \beta_0 - \beta_1 X_{t-1}}$: prior disequilibrium error

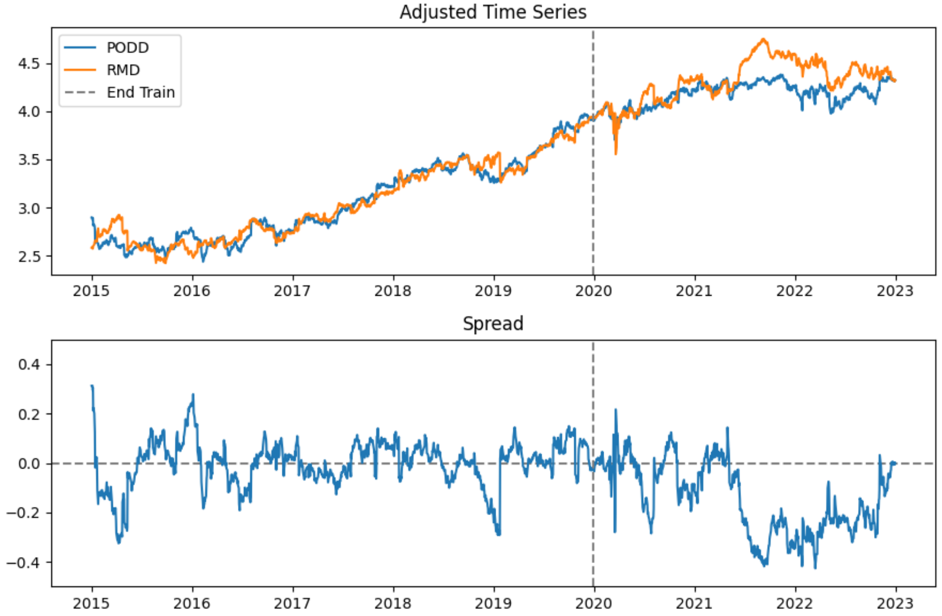
$\boldsymbol{\epsilon_t}$: random shock

$\boldsymbol{\beta_1}$: hedge ratio

$\boldsymbol{\beta_0}$: intercept of long-run equilibrium

- ECM links short-term fluctuations to long-term equilibrium
- If the spread widens, α drives reversion toward the equilibrium level
- In this project, α (intercept) and β are estimated via static OLS, assuming the residuals (spread) follow mean-reverting behavior.
- Model assumption of weakly exogeneous variable X likely to cause misspecification.

Modelling Results:



- Spread (mean-adjusted) is expected to revert towards zero
- AR(1) half-life of residuals in training sample $t_{1/2} = 17.62$ periods and lag coefficient $\rho = 0.9614$, which indicates weak mean reversion
- Mean-reversion characteristic does not hold well in validation period

Cointegration Modelling

Vector Error Correction Model

Vector Error Correction Model (VECM)

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + \mu + \epsilon_t$$

where

$\mathbf{Y_t}$: log price vector for assets

$\mathbf{\Pi = \alpha\beta'}$: long-run equilibrium term

$\boldsymbol{\beta}$: cointegrating vectors of rank r

$\boldsymbol{\alpha}$: adjustment speed

$\mathbf{\Gamma_i}$: short-term lagged adjustments

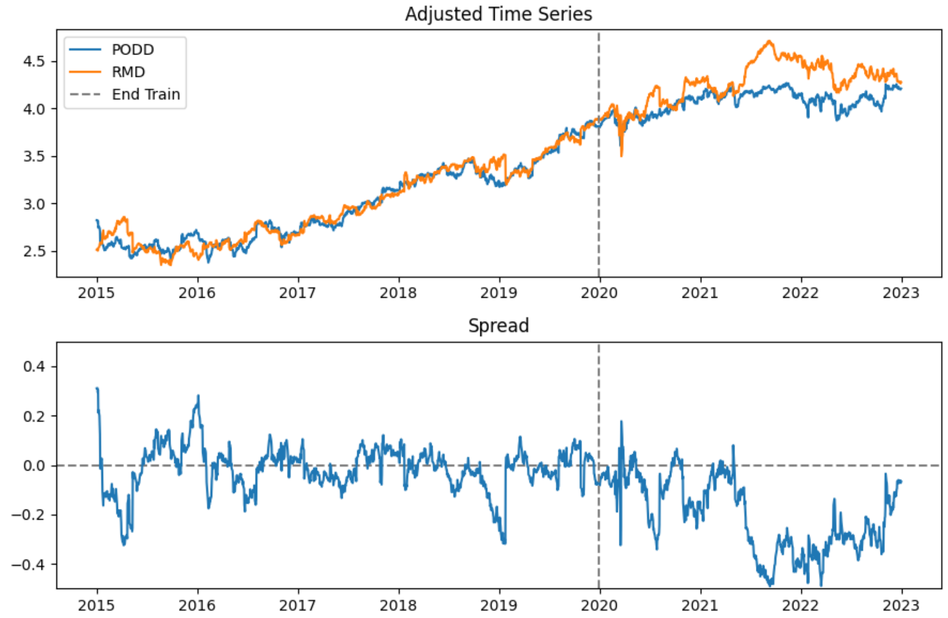
$\mathbf{k}$: lag order (chosen by BIC via VAR)

$\boldsymbol{\mu}$: constant term

$\boldsymbol{\epsilon_t}$: random shock

- The VECM generalizes ECM to model multiple cointegrated assets jointly, capturing both short-term dynamics and long-run equilibrium relationships.
- It avoids directional bias from OLS and provides system-wide correction effects.
- Theoretically, the VECM is supposed to provide a more robust estimate of the cointegrating relationship over the ECM.

Modelling Results:



- $t_{1/2} = 17.3$ periods and $\rho = 0.9606$ suggests minimal improvement
- As before, the mean-reverting behavior only holds for a short period before diverging
- We can consider time-varying models to better capture the long-term relationship

Cointegration Modelling

Rolling Window ECM and VECM

The Rolling ECM/VECM Approach

To predict cointegrating relationship at time t, fit an ECM/VECM on the past n observations $Y = \{y_{t-1}, y_{t-2}, \dots y_{t-n}\}$

Benefits:

- **Time-Varying Parameters:** A rolling model incorporates the time-varying nature of market relationships by allowing following key parameters to evolve – (i) the cointegration relationship, and (ii) the speed of adjustment

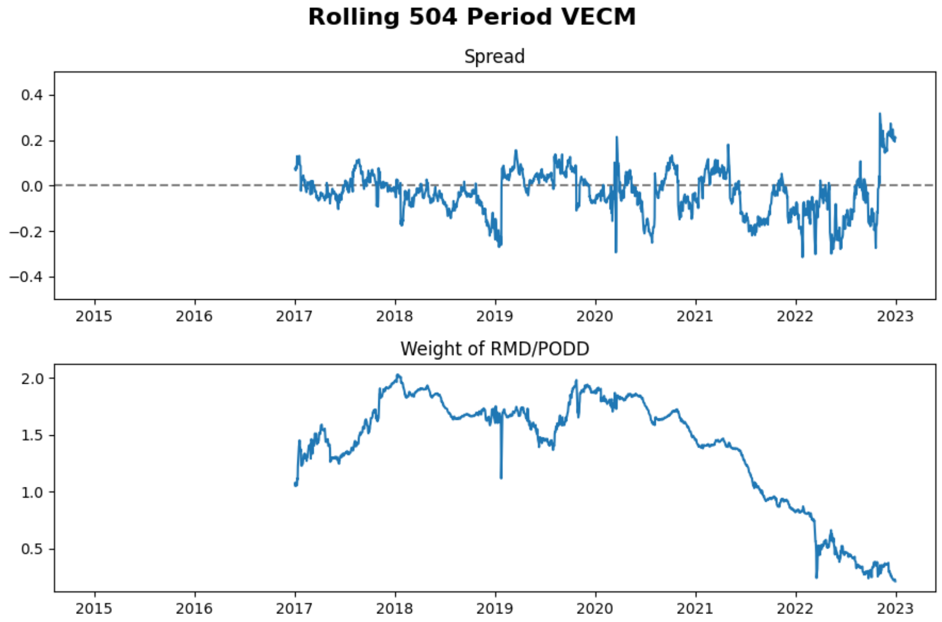
- **Updated Information:** As compared to static approaches, the rolling window continuously incorporates new information and discards potentially outdated information.

Challenges

- **Parameter Choice:** The model suffers from high estimation noise in smaller windows, leading to possibly erratic shifts in estimated parameters. Larger windows on the other hand may be slow to react to genuine structural breaks.

- **Computationally Intensive:** For higher-frequency trading, re-estimation of the VECM can be slow, especially if the step to re-check for evidence of cointegration is done for robustness.

Modelling Results:



- Recalculating spread for each time t based on newly estimated VECM yields $t_{1/2} = 12.5$ periods and $\rho = 0.9459$

- More dangerously, the relative weights are **unstable**, which potentially suggests that a stable long-run relationship does not exist or is poorly estimated. We need a more stable approach.

Kalman Filter State Space Model

1) Predict: $\hat{\boldsymbol{x}}_{t|t-1} = \boldsymbol{F}_t \cdot \hat{\boldsymbol{x}}_{t-1|t-1} + \boldsymbol{B}_t\boldsymbol{u}_t + \boldsymbol{\epsilon}_t$

2) Update: $\hat{\boldsymbol{x}}_{t|t} = \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{K}_t \cdot \boldsymbol{y}_t$

where

$\hat{\mathbf{x}}_{\mathbf{t|t-1}}$: state at t with info till $t-1$

$\mathbf{F}_\mathrm{t}$: state transition matrix, set to $\mathbf{I}$

$\mathbf{B}_\mathrm{t}\mathbf{u}_\mathrm{t}$: deterministic inputs, set to 0

$\boldsymbol{\epsilon}_\mathrm{t}$: process noise term ~ N(0, Q)

$\mathbf{K}_\mathrm{t}$: Kalman gain

$\mathbf{y}_\mathrm{t}$: innovation (prediction residual)

$\mathbf{P}_\mathrm{t}$: state covariance (uncertainty)

$\mathbf{Q}$: process noise covariance

$\mathbf{z}_\mathrm{t}$: observed log price of endog

$\mathbf{H}_\mathrm{t}$: observed log price of exog

$\mathbf{S}_\mathrm{t}$: covariance of innovation

$\mathbf{R}$: measurement noise uncertainty

Intermediate steps between step 1 and 2:

a. Project error covariance: $\boldsymbol{P}_{t|t-1} = \boldsymbol{F}\boldsymbol{P}_{t-1|t-1}\boldsymbol{F}' + \boldsymbol{Q}$

b. Calculate innovation: $\boldsymbol{y}_t = \boldsymbol{z}_t - \boldsymbol{H}_t\boldsymbol{x}_{t|t-1}$

c. Calculate innovation covariance: $\boldsymbol{S}_t = \boldsymbol{H}_t\boldsymbol{P}_{t|t-1}\boldsymbol{H}_t' + \boldsymbol{R}$

d. Calculate Kalman gain: $\boldsymbol{K}_t = \boldsymbol{P}_{t|t-1}\boldsymbol{H}_t'\boldsymbol{S}_t^{-1} + \boldsymbol{R}$

e. Update error covariance: $\boldsymbol{P}_{t|t} = (\boldsymbol{I} - \boldsymbol{K}_t\boldsymbol{H}_t)\boldsymbol{P}_{t|t-1}$

Cointegration Modelling

Kalman Filter in Use

Kalman Filter

1) Predict: $\hat{\boldsymbol{x}}_{t|t-1} = \boldsymbol{F}_t \cdot \hat{\boldsymbol{x}}_{t-1|t-1} + \boldsymbol{B}_t \boldsymbol{u}_t + \boldsymbol{\epsilon}_t$

2) Update: $\hat{\boldsymbol{x}}_{t|t} = \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{K}_t \cdot \boldsymbol{y}_t$

- Recursively predicts the evolution of a state and adaptively adjusts the model using new info to improve future predictions
- In our model, we initialize the initial state as the VECM estimate

Benefits:

- **Responsiveness:** Filter incorporates new information and incorporates it with a weightage based on the calculated Kalman gain.
- **Low-Cost:** The model is highly efficient and its recursive nature means it has fast compute speed and reduced memory requirement

Challenges

- **Tuning Complexity:** To balance measurements and predictions, the model needs the process noise covariance matrix $\boldsymbol{Q}$ and measurement noise uncertainty $\boldsymbol{R}$ as inputs, which are not known precisely

Modelling Results:

- Significant improvement of $t_{1/2} = 4.3$ periods and $\rho = 0.8518$
- Relative weights are more consistent and aligns with our underlying assumption of a long-run cointegration relationship

Interpretation of Model Weights

Transforming Model Outputs to a Trading Strategy

Trading Weights

- **Intuition:** Using the ECM/VECM model, or an extension (Kalman Filter), we estimate the long-run cointegrating relationship and trade deviations from it (i.e., the spread).
- By treating the PODD/RDM spread as a ***singular asset*** $\boldsymbol{u_t}$, we take positions based on its relative value around its mean.
- From the cointegrating relationship, we know the relative (dollar) weights of each asset to replicate u_t.
- **Normalization**: We ***normalize*** u_t such it represents 100% of the purchasing power of our portfolio, i.e., $u_t^* \in [0, 1]$.
  - Under US Reg T margin, one can borrow up to 50% (initial margin) of a long position and must hold at least 50% (initial margin) of a short position in cash in their account.

$$\boxed{u_t^* = \frac{u_t}{\frac{|\text{Short Weights}|}{50\%} + \frac{|\text{Long Weights}|}{50\%}} = 2}$$

Trading Signal

- **Intuition:** The signal θ_t informs us of (i) ***when to long or short the spread***, and (ii) the ***sizing*** of the position.
- We bound $\theta_t \in [-1, 1]$ where both boundary values intuitively represent a 100% long/short respectively.
- **Standardization:** The signal is constructed as a function of the n-rolling z-score of the spread, providing a standardized interpretation of the deviation of the spread from its mean.

$$z_t = \frac{u_t - \bar{u}_{t-1,\dots,t-n}}{\sigma_{u_{t-1,\dots,t-n}}}$$

- Depending on the strategy, there are **2 ways** to construct the signal:
  - 1. Threshold Strategy (All-in or All-out)
  - 2. Linear Strategy (Proportional sizing)

Defining The Trading Strategy

Constructing trading signals based on observed spreads

Threshold Strategy

- **Intuition:** Go all-in or all-out depending on a ***predefined absolute cutoff***.

- **Entry/Exit:**

$$\theta_t\left(z_t, \theta_{t-1}\,; k_{entry}, k_{exit}\right) = \begin{cases} 1, & if\ z_t \le -k_{entry} \\ 1, & if\ -k_{entry} < z_t < -k_{exit}\ and\ \theta_{t-1} = 1 \\ 0, & if\ -k_{exit} \le z_t < k_{entry}\ and\ \theta_{t-1} = 1 \\ -1, & if\ z_t \ge k_{entry} \\ -1, & if\ k_{exit} \le z_t < k_{entry}\ and\ \theta_{t-1} = -1 \\ 0, & if\ -k_{entry} < z_t \le k_{exit}\ and\ \theta_{t-1} = -1 \end{cases}$$

- **Pros/Cons:**

(+) Simple. Specifically target high-conviction extreme events, avoiding transaction costs on weak signals.

(+) Lower sensitivity from disregarding magnitude of z-score reduces participation rate.

(-) Highly dependent on the selected threshold, which may not hold in practice.

Linear Strategy

- **Intuition:** Scale position size ***according to magnitude of the z-score***, eventually capping off at a predefined cutoff.

- **Entry/Exit:**

$$\theta_t(z_t\,; k_{threshold}) = \left[\frac{z_t}{k_{threshold}}\right]_{-1}^{+1}$$

- **Pros/Cons:**

(+) Proportional response to z-score better utilizes available information for its decision.

(+) Higher sensitivity allows gradual scaling into a position, minimizing dependence around a threshold.

(-) Sensitivity to noise will lead to slippage from trading weak signals.

A balanced approach may consider the combination of both strategies.

Bayesian Optimisation

Optimising Hyperparameters by Maximising Sharpe Ratio

Bayesian Optimization is a sequential model-based approach for **optimizing black-box functions (functions that are expensive to evaluate and have no analytical form).** It uses **probabilistic models** to predict which parameters are most likely to **yield the best results**.

Objective Function

Maximize Sharpe Ratio

Parameters

1. Entry Threshold
2. Exit Threshold
3. Z –score rolling window
4. Model Specific Parameters

Framework

1. Define parameters bounds
2. Apply Bayesian Optimization using gp_minimize() to find best parameters

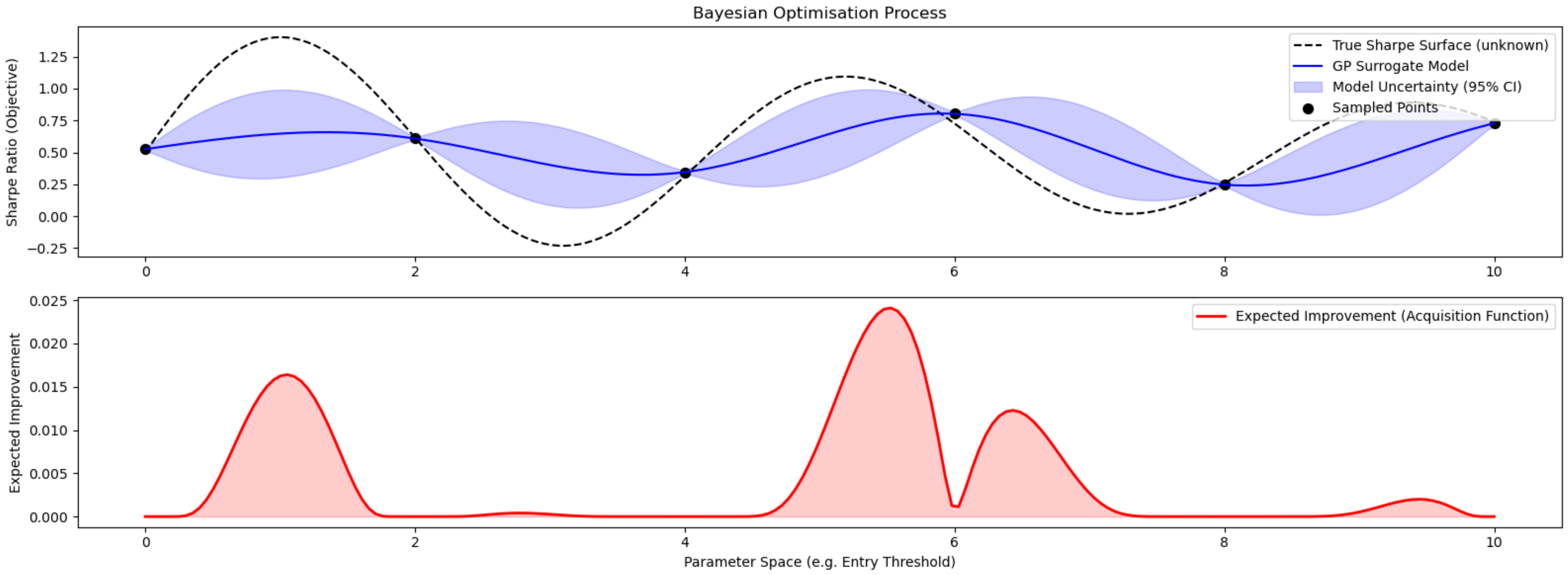
Parameter	Static OLS	Rolling OLS	Static VECM	Rolling VECM	Kalman Filter	Range
Entry Threshold	✅	✅	✅	✅	✅	±1.5 – ±2.5
Exit Threshold	✅	✅	✅	✅	✅	±0.2 – ±1.0
Z-score Window	✅	✅	✅	✅	✅	20 – 120
Model Lookback Window		✅		✅		200 – 800 days
Kalman Qβ (Process Noise)					✅	$1e^{-5} – 1e^{-2}$
Kalman Qintercept					✅	$1e^{-7} – 1e^{-4}$
Kalman R (Measurement Noise)					✅	$1e^{-4} – 1$

Bayesian Optimisation

How the Process works

SMU SINGAPORE MANAGEMENT UNIVERSITY

1. **Evaluate Initial Points**
   - The optimizer begins by evaluating a set of randomly chosen parameter combinations to get an initial understanding of the parameter space
2. **Construct the Surrogate Model**
   - Bayesian Optimization constructs a probabilistic model (usually a Gaussian Process) based on the initial evaluations
3. **Exploitation & Exploration**
   - Using an acquisition function (like Expected Improvement), the optimizer balances between exploitation (trying parameter values near previously successful results) and exploration (trying new, untested areas of the parameter space)
4. **Update the Model and Repeat**
   - The optimizer evaluates the suggested parameter set, updates the surrogate model with the new result, and repeats the process for the remaining iterations
   - The process continues until it finds the parameter combination that maximizes the Sharpe ratio within the given bounds



Backtesting

Setting Up the Backtest

Simulation Parameters

- **Validation Period:** 1 Jan 2020 to 31 Dec 2022 – for model tuning
- **Trading Period:** 1 Jan 2023 to 31 Oct 2025, Daily Close

Interest/Costs	Value	Description
Risk Free Rate (RFR)	4.00%	Reference benchmark rate for interest and cost of borrowing
Short Borrowing Fee	0.25%	Fee charged as % of short value; for borrowing stocks, assume stock is liquid
Transaction Costs	0.20%	Fees charged as % of transaction value; includes bid-ask spread, broker comms, etc.
Margin Borrowing Fee	RFR	Fees charges as % of borrowed margin; for positions exceeding portfolio equity
Interest on Idle Cash	RFR	Interest earned as % of idle cash; for returns on unused capital
Interest on Short Collateral	RFR	Interest earned as % of collateral posted on short positions; for returns on short value

Backtest Metrics

- **Annualized Excess Return:** Measures profit capacity of trading strategy, excess of RFR

$$Ann. Return = Daily\ Return\ \times \sqrt{252}\ - RFR$$

- **Sharpe Ratio:** Widely-used relative risk-reward metric
- **Max Drawdown (MDD):** Measures maximum pain
- **Longest Drawdown Duration:** Opportunity cost
- **Calmar Ratio:** Measures relative reward for enduring loss.

$$\frac{Ann.\ Excess\ Returns}{MDD}$$

- **Sortino Ratio:** Measures reward only against harmful risk.

$$\frac{Ann.\ Excess\ Returns}{Downside\ Volatility}$$

Backtest Results

Out of sample testing results

Model	Trading Strategy	Ann. Excess Return	MDD	LDD (periods)	Sharpe Ratio	Calmar Ratio	Sortino Ratio
Static ECM	Threshold 2 / 0.5	39.79%	-20.33%	86	1.5287	1.9574	2.5075
Static VECM	Threshold 2 / 0.5	37.64%	-20.20%	86	1.5160	1.8632	2.4375
Rolling 504-period ECM	Threshold 2 / 0.5	-3.25%	-63.65%	561	-0.1237	-0.0510	-0.1450
Rolling 504-period VECM	Threshold 2 / 0.5	-14.82%	-89.84%	660	-0.5661	-0.1650	-0.5649
Kalman Filter	Threshold 2 / 0.5	22.00%	-16.46%	106	0.9826	1.3364	1.6915
Kalman Filter + Bayesian Optimized Q and R	Threshold 2 / 0.5	33.44%	-12.36%	53	1.3200	2.7043	2.1468
Kalman Filter + Bayesian Optimized Q, R, entry/exit, z	Threshold 2.3 / 0.8	42.16%	-18.31%	73	1.4955	2.3033	2.4112

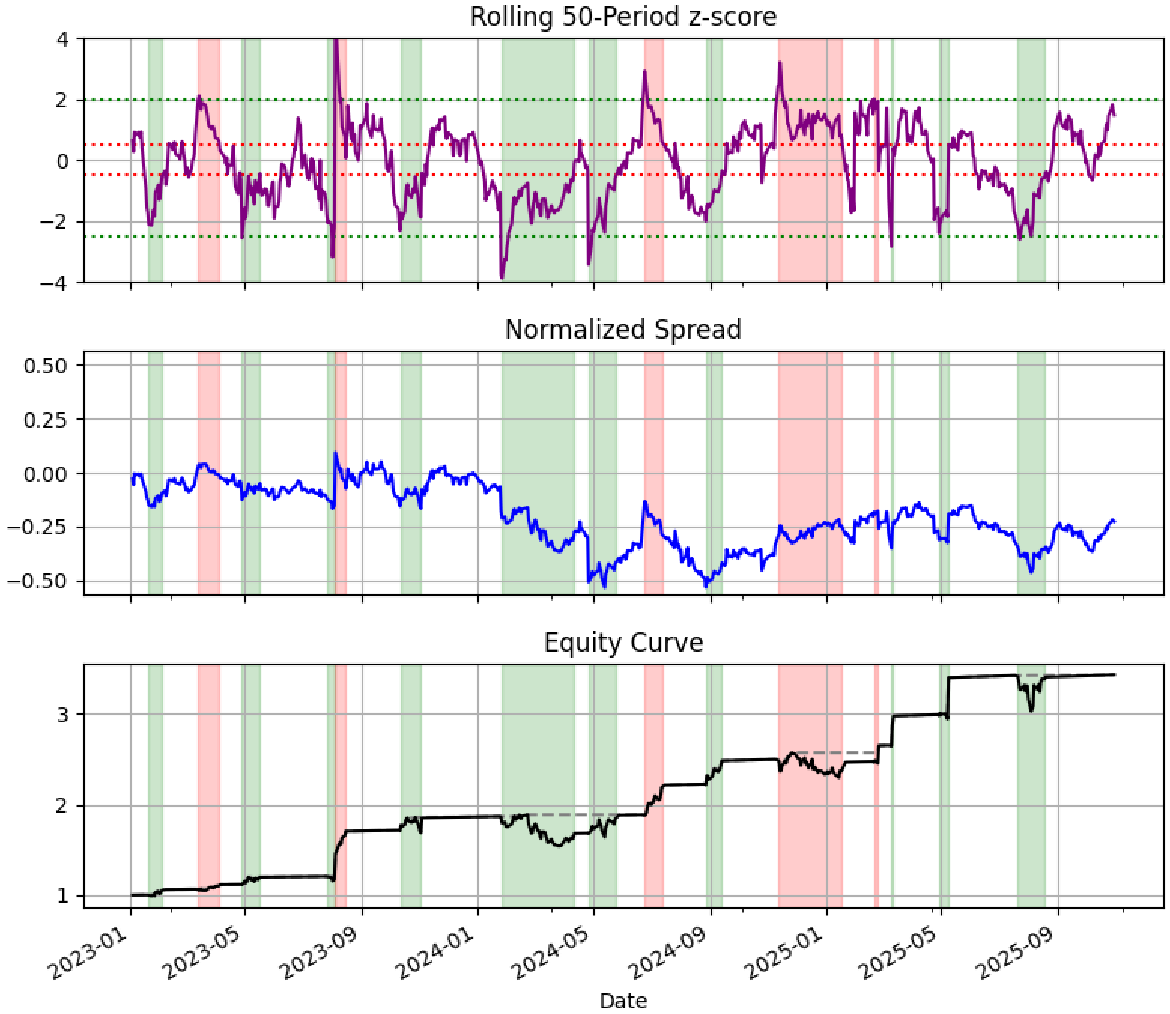
Observations:

- Kalman Filter after optimization suggests stronger results in terms of higher returns with more controlled drawdowns
- Rolling models fail heavily, as expected from the large variation in modelled weights
- Linear strategies mostly underperformed against threshold strategies, likely due to high slippage

Backtest Results

Static ECM Model

Static ECM Model



Trade Parameters
• Z-score window: 100 periods • Enter / Exit Thresholds: 2.0 / 0.5

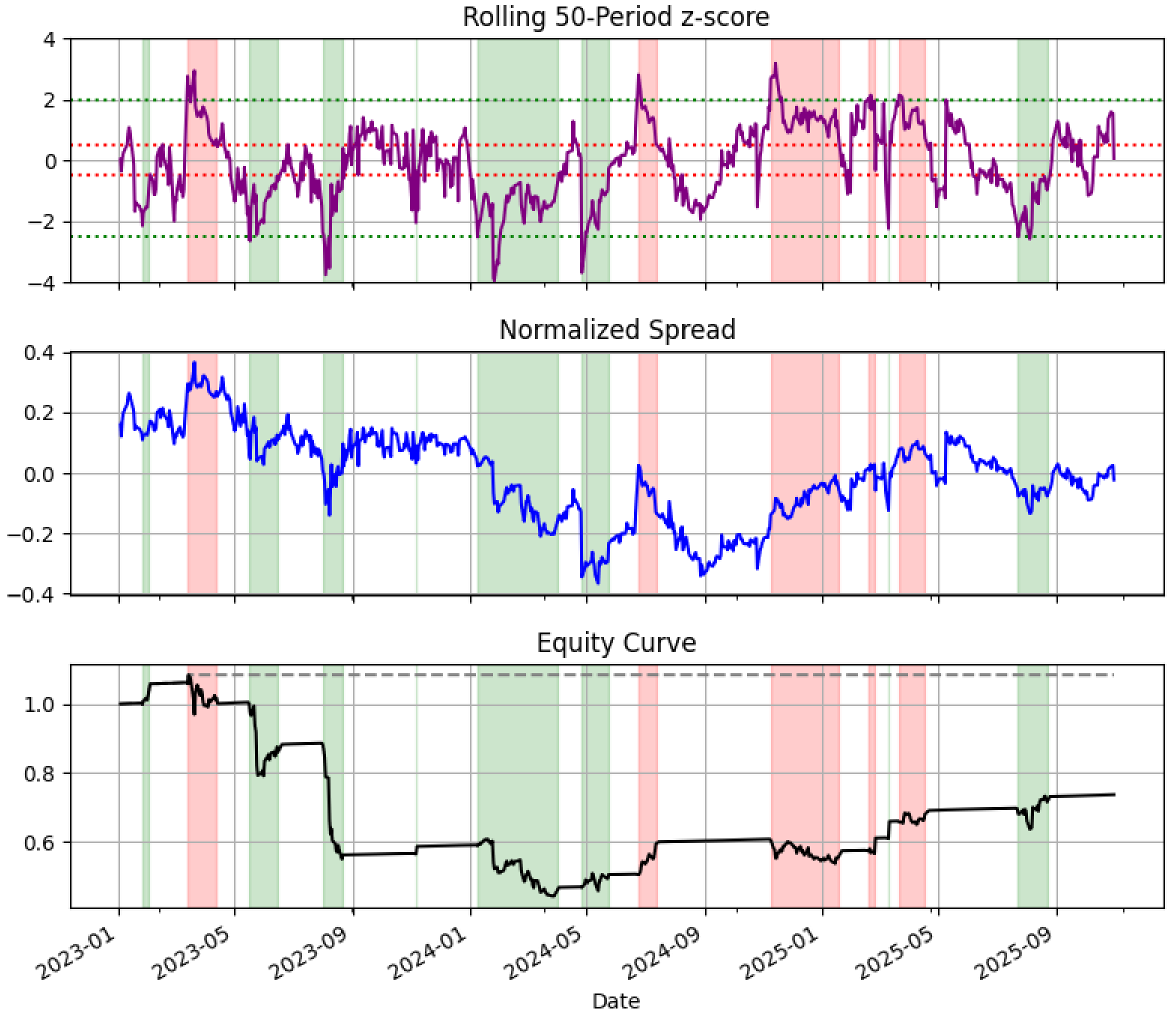
Annualized Excess Return	39.79%
Max Drawdown (MDD)	-20.33%
Longest Drawdown Duration	86 periods
Sharpe Ratio	1.5287
Calmar Ratio	1.9574
Sortino Ratio	2.5075

- Despite mean-reverting relationship being poorly modelled, the short z-score window allowed the model to still react to the shifting short-term mean and profit of reversion to it.
- This outcome, though favorable, is not intended as a predicted behavior of our model. The same applies to VECM

Backtest Results

Rolling VECM Model

Rolling VECM Model



Trade Parameters
• Rolling window: 504 periods (2 years) • Z-score window: 100 periods • Enter / Exit Thresholds: 2.0 / 0.5

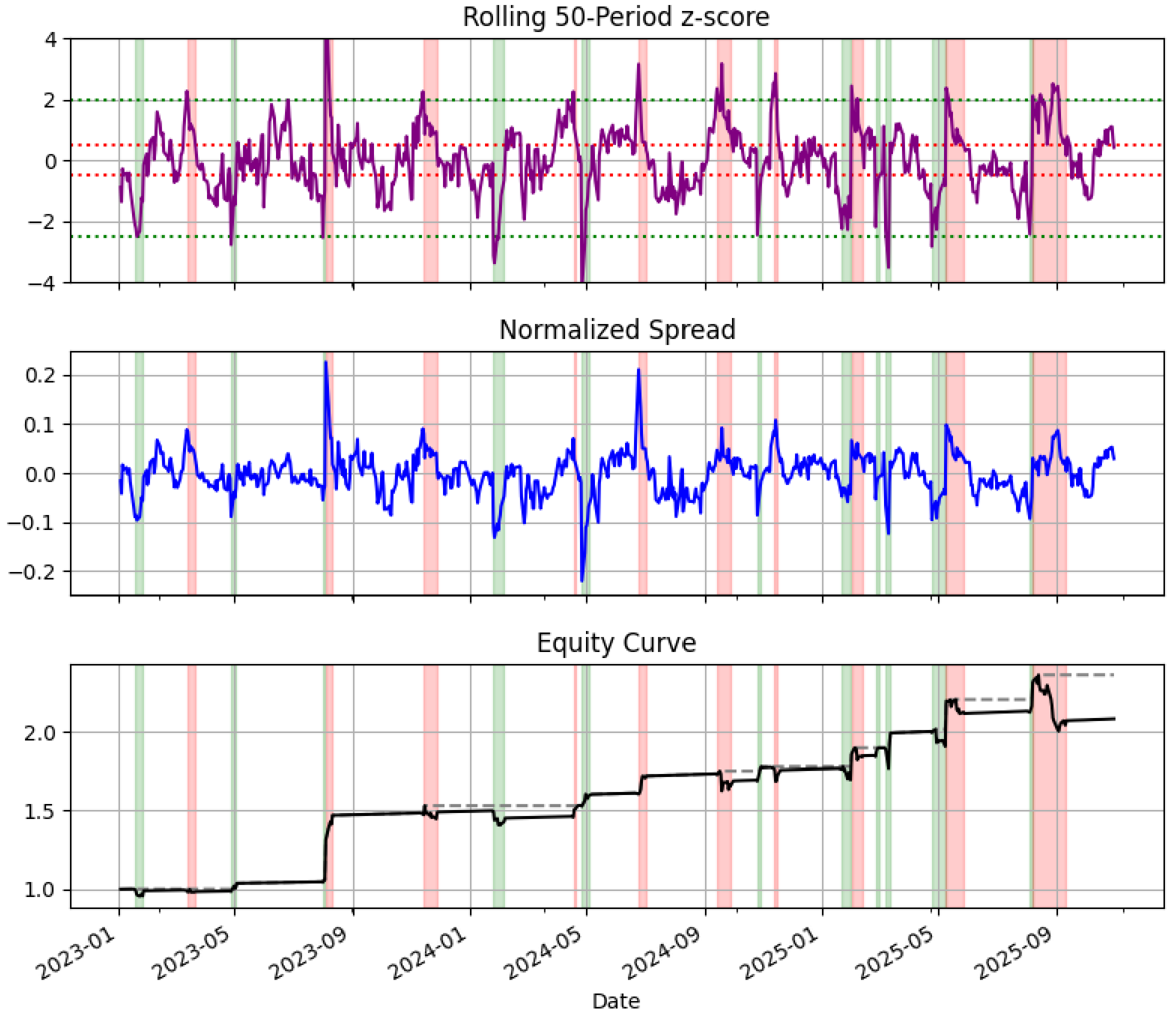
Annualized Excess Return	-14.82%
Max Drawdown (MDD)	-89.84%
Longest Drawdown Duration	660 periods
Sharpe Ratio	-0.5661
Calmar Ratio	-0.1650
Sortino Ratio	-0.5649

- The model's failure to maintain a stable cointegrating relationship led to erosion of profits from the rebalancing of weights.

Backtest Results

Kalman Filter Model

Kalman Filter Model



Trade Parameters		
$\mathbf{P} = 10^{-5} \cdot \mathbf{I}$	$\mathbf{Q}_{\beta} = 10^{-3}$ $\mathbf{Q}_{\mu} = 10^{-6}$	$\mathbf{R} = \mathbf{I}$
Z-score window: 100 periods	Enter / Exit Thresholds: 2.0 / 0.5	

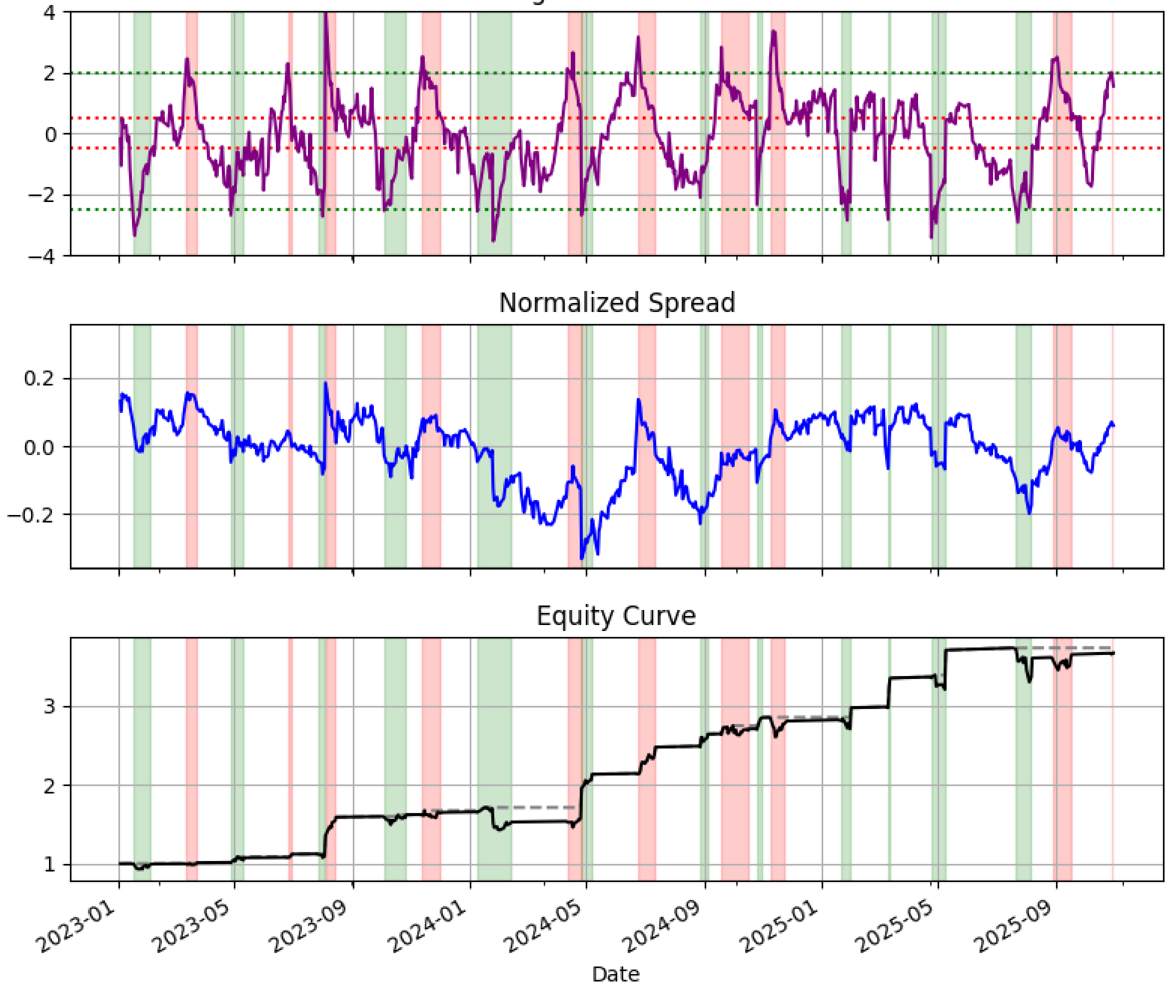
Annualized Excess Return	22.00%
Max Drawdown (MDD)	-16.46%
Longest Drawdown Duration	106 periods
Sharpe Ratio	0.9826
Calmar Ratio	1.3200
Sortino Ratio	1.4955

- The Kalman Filter models mean-reversion of the cointegrating pair in a stable manner, allowing it to accurately identify divergences from its long-run relationship.
- Steady P&L is achieved with relatively low drawdowns.

Backtest Results

Bayesian Optimized Kalman Filter Model

Bayesian Optimized Kalman Filter Model



Trade Parameters		
$\mathbf{P} = 10^{-5} \cdot \mathbf{I}$	$\mathbf{Q}_\beta = 3.08 \cdot 10^{-5}$ $\mathbf{Q}_\mu = 10^{-8}$	$\mathbf{R} = 5 \cdot \mathbf{I}$
Z-score window: 32 periods	Enter / Exit Thresholds: 2.3 / 0.8	

Annualized Excess Return	42.16%
Max Drawdown (MDD)	-18.31%
Longest Drawdown Duration	73 periods
Sharpe Ratio	1.4955
Calmar Ratio	2.3033
Sortino Ratio	2.4112

- Further optimization of Kalman Filter's parameters improves excess returns and Sharpe ratio
- Stability of model makes it highly viable for use

Reflections

Challenges/Limitations, Potential Improvements, Main Takeaways

SMU SINGAPORE MANAGEMENT UNIVERSITY

Challenges/Limitations

- **High computational cost** from rolling estimations and Bayesian optimization
- **Parameter sensitivity** where performance depends on window length and lag choice
- Data constraints where short windows add noise and long windows reduce responsiveness

Assumptions

- **Cointegration relationships remain stable** during training and test periods
- Parameters optimized once are assumed constant in out-of-sample data

Key Takeaways

- Used **economic similarity filters** and **cointegration tests** to select stable and meaningful pairs
- Compared **different econometric models** to capture mean-reverting dynamics
- Applied **Bayesian optimization** to enhance Sharpe ratio and parameter robustness
- Emphasized the **importance of validation** to prevent overfitting
- Built an end-to-end quantitative trading pipeline from data cleaning to backtesting

Potential Improvements

- Implement a **rolling train–validate–test framework** to track performance over time
- Explore regime-switching or machine learning models for adaptive behavior
- Conduct **sensitivity analysis** to test parameter robustness under varying conditions
- Benchmark Bayesian optimization against alternative methods such as grid or genetic search
- Extend from pair-level to portfolio-level optimization

Thank You!