①

$$Z_t = X_t - \hat{\beta} Y_t$$

↓ Residuals ↓ cointegration coefficient (from OLS)

If Z_t is stationary (The ADF test rejects the null hypothesis)

⇒ series are cointegrated.

$$Z\text{-Score}_t = \frac{Z_t - \mu_Z}{\sigma_Z}$$

← spread's rolling mean

σ_Z ← standard deviation

Z-score > +Z-threshold (θ_{in}): sell signal
short the overvalued
buy the undervalued

Z-score < −Z-threshold ($-\theta_{in}$): buy signal
long the underpriced
short the overpriced

|Z-score| < θ_{out} : close position.

In the code: $\theta_{in} = 1.5$, $\theta_{out} = 0.5$
Enter Exit

② If θ_{in} is set too high: the model wait for extremely pronounced divergences

If θ_{in} is too low: may engage in frequent, low-quality trades

If θ_{out} too close to 0: may exit trades prematurely

If higher θ_{out}: increasing the risk of missed reversal singal or persistent losses

③ How to optimize parameters?

Training Period	Validation Period	Test Period

↘ Optional, preventing overfitting to the train set.

Objective function: maximizing cumulative returns over the Training period.

Training period

Grid Search

$\theta_{in} \in \{1.0, 1.1, \cdots, 2.5\}$
$\theta_{out} \in \{0.0, 0.1, \cdots, 1.0\}$

$16 \times 11 = 176$

Check which combination of θ_{in} and θ_{out} has the highest cumulative profit.

Bayesian Optimization (Optuna)

① Try a few points randomly

$(\theta_{in}$	$\theta_{out})$	Return	
(1.8,	0.6)	5%	
(1.2,	0.3)	8%	✓
(2.2,	0.8)	3%	

It seems lower θ_{in} and θ_{out} perform better
Next focus on the region
(1.0~1.5, 0.1~0.5)

② (1.3, 0.4) 8.5%
(1.1, 0.4) 9.2% ✓

Abandon the poorly performing regions

③ (1.5, 0.6)
Occasionally try other regions

(1.0, 0.1)
Try boundary testing.

Validation Period

If these parameters also produce favorable results in the validation periods, this indicates robustness, the chosen thresholds are not simply an artifact of overfitting to the past, and Generalized well.

---

④ However, parameter optimization is not always the best option. The example in paper obtained the same cumulative return (5.2%) using optimized parameters ($\theta_{in}=1.42$, $\theta_{out}=0.37$) and fixed parameters ($\theta_{in}=2$, $\theta_{out}=1$). However, the other statistics were higher.

In conclusion, maybe choose not to optimize? (3.4)