Pairs Trading

Prof. Daniel P. Palomar
The Hong Kong University of Science and Technology (HKUST)

MAFS6010R- Portfolio Optimization with R
MSc in Financial Mathematics
Fall 2018-19, HKUST, Hong Kong
Outline

1. Cointegration
2. Basic Idea of Pairs Trading
3. Design of Pairs Trading
   - Pairs selection
   - Cointegration test
   - Optimum threshold
4. LS Regression and Kalman for Pairs Trading
5. From Pairs Trading to Statistical Arbitrage (StatArb)
   - VECM
   - Factor models
   - Optimization of mean-reverting portfolio (MRP)
6. Summary
Outline

1. **Cointegration**

2. **Basic Idea of Pairs Trading**

3. **Design of Pairs Trading**
   - Pairs selection
   - Cointegration test
   - Optimum threshold

4. **LS Regression and Kalman for Pairs Trading**

5. **From Pairs Trading to Statistical Arbitrage (StatArb)**
   - VECM
   - Factor models
   - Optimization of mean-reverting portfolio (MRP)

6. **Summary**
Cointegration

- Cointegration is a very interesting property that can be exploited in finance for trading.
- Idea: While it may be difficult to predict individual stocks, it may be easier to predict relative behavior of stocks.
- Illustrative example: A drunk man is wandering the streets (random walk) with a dog. Both paths of man and dog are nonstationary and difficult to predict, but the distance between them is mean-reverting and stationary.
Correlation vs. cointegration

- Everybody is familiar with the concept of correlation between two random variables:
  - correlation is high when they co-move
  - correlation is zero when they move independently

- So what is cointegration?
  - cointegration is high when two quantities move together or remain close to each other
  - cointegration is inexistent if the two quantities do not stay together

- Clear? 😊 You can see why this concept may be difficult to grasp at first, but the truth is that it’s easy.¹

- In the financial context:
  - Cointegration of (log-)prices $y_t$ refers to long-term co-movements.
  - Correlation of (log-)returns $\Delta y_t = y_t - y_{t-1}$ characterizes short-term co-movements in (log-)prices $y_t$.

---

Correlation vs. cointegration

- Example of high correlation with no cointegration:
Indeed the returns are highly correlated, see scatter plot:
Opposite example of high cointegration with no correlation:
Indeed the returns are not correlated, see scatter plot:
A time series is called integrated of order $p$, denoted as $I(p)$, if the time series obtained
- by differencing the time series $p$ times is weakly stationary,
- while by differencing the time series $p - 1$ times is not weakly stationary.

**Example:** stock log-prices $y_t$ are integrated of order $I(1)$ because
- log-prices are not stationary
- but log-returns $y_t - y_{t-1}$ are stationary (at least for some period of time).

A multivariate time series is said to be cointegrated if it has at least one linear combination being integrated of a lower order, e.g., $y_t$ is not stationary but $w^T y_t$ is stationary for some weights $w$. 
Cointegration

- Consider the following two nonstationary time series (e.g., log-prices of stocks):

\[ y_{1t} = \gamma x_t + w_{1t} \]
\[ y_{2t} = x_t + w_{2t} \]

with a stochastic common trend defined as a random walk:

\[ x_t = x_{t-1} + w_t \]

where \( w_{1t}, w_{2t}, w_t \) are i.i.d. residual terms mutually independent.

- The coefficient \( \gamma \) is the secret ingredient here.
- If \( \gamma \) is known, then we can define the so-called “spread”

\[ z_t = y_{1t} - \gamma y_{2t} = w_{1t} - \gamma w_{2t} \]

which is stationary and \textbf{mean reverting}.

- Interestingly, the differences (i.e., log-returns) \( \Delta y_{1t} \) and \( \Delta y_{2t} \) can have an arbitrarily small correlation:

\[ \rho = 1/\left( \sqrt{1 + 2\sigma_1^2/\sigma^2} \sqrt{1 + 2\sigma_2^2/\sigma^2} \right). \]
Cointegration

- The log-prices $y_{1t}$ and $y_{2t}$ are cointegrated and the spread $z_t = y_{1t} - \gamma y_{2t}$ is stationary (assume $\gamma = 1$):
Cointegration

2 Basic Idea of Pairs Trading

3 Design of Pairs Trading
   - Pairs selection
   - Cointegration test
   - Optimum threshold

4 LS Regression and Kalman for Pairs Trading

5 From Pairs Trading to Statistical Arbitrage (StatArb)
   - VECM
   - Factor models
   - Optimization of mean-reverting portfolio (MRP)

6 Summary
Basic Idea of Pairs Trading

- Recall that if two time series are cointegrated, then in the long term they remain close to each other.
- In other words, the spread $z_t = y_{1t} - \gamma y_{2t}$ is mean reverting.
- This mean-reverting property of the spread can be exploited for trading and it is commonly referred to as “pairs trading” or “statistical arbitrage”.
- The idea behind pairs trading is to
  - short-sell the relatively overvalued stocks and buy the relatively undervalued stocks,
  - unwind the position when they are relatively fairly valued.
Trading the spread

- Suppose the spread $z_t = y_{1t} - \gamma y_{2t}$ is mean-reverting with zero mean.

- Stat-arb trading:
  - if spread is low ($z_t < -s_0$), then stock 1 is undervalued and stock 2 overvalued:
    - buy the spread (i.e., buy stock 1 and short-sell stock 2)
    - unwind the positions when it reverts to zero after $i$ time steps $z_{t+i} = 0$
  - if spread is high ($z_t > s_0$), then stock 1 is overvalued and stock 2 undervalued:
    - short-sell the spread (i.e., short-sell stock 1 and buy stock 2)
    - unwind the positions when it reverts to zero after $i$ time steps $z_{t+i} = 0$

- The profit, say, from buying low and unwinding at zero is $z_{t+i} - z_t = s_0$. So easy!

- Indeed $z_{t+i} - z_t = -\gamma(y_{2,t+i} - y_{2t}) + (y_{1,t+i} - y_{1t})$, so the whole process is like having used a portfolio with weights $w = \begin{bmatrix} 1 \\ -\gamma \end{bmatrix}$.

- Recall that the return of a portfolio $w$ is $w^T \Delta y_t$. 
Trading the spread

Illustration on how to trade the spread $z_t = y_{1t} - \gamma y_{2t}$:

Pairs trading or statistical arbitrage

- Statistical arbitrage can be used in practice with profits:

\[\text{Spread}\]

\[\text{Position}\]

\[\text{P&L}\]

But how to discover cointegrated pairs and $\gamma$?

- One interesting approach is based on a VECM modeling of the universe of stocks: From the parameter $\beta$ contained in the low-rank matrix $\mathbf{\Pi} = \alpha \beta^T$ one can extract a cointegration subspace. After that, one can design some portfolio within that cointegration subspace.\(^4\)

- A simpler approach to discover pairs is by brute force, i.e., try exhaustively different combinations of pairs of stocks and see if they are cointegrated.

- But, given a potential pair, how do we obtain the “secret” $\gamma$?

- Easy! Just a simple LS regression!

Recall that

- $\gamma$ is needed to form the spread to be traded (i.e., portfolio)
- the spread mean $\mu$ is needed to determine the thresholds for entering a trade and unwind later the position.

Outline

1. Cointegration
2. Basic Idea of Pairs Trading
3. Design of Pairs Trading
   - Pairs selection
   - Cointegration test
   - Optimum threshold
4. LS Regression and Kalman for Pairs Trading
5. From Pairs Trading to Statistical Arbitrage (StatArb)
   - VECM
   - Factor models
   - Optimization of mean-reverting portfolio (MRP)
6. Summary
Design of a pairs trading strategy

We first focus on pairs trading (i.e., statistical arbitrage between two stocks) as the example to introduce the main steps of statistical arbitrage.

In practice, pairs trading contains three main steps\(^5\):

- **Pairs selection**: identify stock pairs that could potentially be cointegrated.
- **Cointegration test**: test whether the identified stock pairs are indeed cointegrated or not.
- **Trading strategy design**: study the spread dynamics and design proper trading rules.

Outline

1. Cointegration
2. Basic Idea of Pairs Trading
3. Design of Pairs Trading
   - Pairs selection
   - Cointegration test
   - Optimum threshold
4. LS Regression and Kalman for Pairs Trading
5. From Pairs Trading to Statistical Arbitrage (StatArb)
   - VECM
   - Factor models
   - Optimization of mean-reverting portfolio (MRP)
6. Summary
Normalized price distance\(^6\) (as a rough proxy to measure cointegration):

\[
\text{NPD} \triangleq \sum_{t=1}^{T} (\tilde{p}_{1t} - \tilde{p}_{2t})^2
\]

where the normalized price \(\tilde{p}_{1t}\) of stock 1 is given by \(\tilde{p}_{1t} = p_{1t}/p_{10}\). The normalized prices of stock 2 defined similarly.

One can easily (i.e., cheaply) compute the NPD for all the possible combination of pairs and select some pairs with smallest NPD as the potentially cointegrated pairs.

Later one can use a more refined measure of cointegration (more computationally demanding).

---

Outline

1. Cointegration
2. Basic Idea of Pairs Trading
3. Design of Pairs Trading
   - Pairs selection
   - Cointegration test
   - Optimum threshold
4. LS Regression and Kalman for Pairs Trading
5. From Pairs Trading to Statistical Arbitrage (StatArb)
   - VECM
   - Factor models
   - Optimization of mean-reverting portfolio (MRP)
6. Summary
Least Squares (LS) regression

- If the spread $z_t$ is stationary, it can be written as

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

where

- $\mu$ represents the equilibrium value and
- $\epsilon_t$ is a zero-mean residual.

- Equivalently, it can be written as

$$y_{1t} = \mu + \gamma y_{2t} + \epsilon_t$$

which now has the typical form of linear regression.

- Least squares (LS) regression over $T$ observations:

$$\minimize_{\mu, \gamma} \sum_{t=1}^{T} (y_{1t} - (\mu + \gamma y_{2t}))^2$$

---

LS regression is used to estimate the parameters $\mu$ and $\gamma$, obtaining the estimates $\hat{\mu}$ and $\hat{\gamma}$.

If $y_{1t}$ and $y_{2t}$ are $I(1)$ and are cointegrated, then the estimates converge to the true values as the number of observations goes to infinity\(^8\).

Using the estimated parameters $\hat{\mu}$ and $\hat{\gamma}$, we can compute the residuals

\[ \hat{\epsilon}_t = y_{1t} - \hat{\gamma}y_{2t} - \hat{\mu}. \]

Then, one has to decide whether the spread is stationary, i.e., $\epsilon_t$ is stationary. In practice, the estimated residuals are used $\hat{\epsilon}_t$.

There are many well-defined mathematical tests for the stationarity of $\hat{\epsilon}_t$, e.g., augmented Dicky-Fuller (ADF) test, Johansen test, etc.

---

Outline

1. Cointegration
2. Basic Idea of Pairs Trading
3. Design of Pairs Trading
   - Pairs selection
   - Cointegration test
   - Optimum threshold
4. LS Regression and Kalman for Pairs Trading
5. From Pairs Trading to Statistical Arbitrage (StatArb)
   - VECM
   - Factor models
   - Optimization of mean-reverting portfolio (MRP)
6. Summary
Optimum threshold

- Once some identified pairs have passed the cointegration test, one still needs to decide the entry and exit thresholds to open and unwind the positions, respectively.
- For the sake of concreteness, we focus on studying the entry threshold:
  - open positions when the spread diverges from its long-term mean by $s_0$
  - unwind the position when it reverts to its mean
- Thus, the key problem now is how to design the value of $s_0$ such that the total profit is maximized.
- Total profit:
  
  \[ \text{profit of each trade} \times \text{number of trades} \]

  - profit of each trade is $s_0$
  - number of trades is related to the zero crossings, which can be analyzed theoretically as well as empirically.
- We focus now on estimating the number of trades.
Optimum threshold $s_0$: Parametric approach

- Suppose the spread follows a standard Normal distribution.
- The probability that the spread deviates above from the mean by $s_0$ or more is
  \[ 1 - \Phi(s_0) \]
  where $\Phi(\cdot)$ is the c.d.f. of the standard Normal distribution.
- For a path with $T$ days, the number of tradable events is
  \[ T(1 - \Phi(s_0)). \]
- For each trade, the profit is $s_0$ and then the total profit is $s_0 T(1 - \Phi(s_0))$.
- Then the optimal threshold is $s_0^* = \arg \max_{s_0} \{ s_0 T(1 - \Phi(s_0)) \}$.
- In practice, one cannot know the true distribution but can estimate the distribution parameters.
- Then one can compute the total profit based on estimated distribution.
Optimum threshold $s_0$: Parametric approach

- Optimal threshold $s_0^*$ maximizes the total profit:

![Graphs showing probability of trades, profit of each trade, and total profit for theoretical and parametric approaches.](image)
Optimum threshold $s_0$: Non-parametric approach

- Suppose the observed sample path has length $T$: $z_1, z_2, \ldots, z_T$.
- We consider $J$ discretized threshold values as $s_0 \in \{s_{01}, s_{02}, \ldots, s_{0J}\}$ and the empirical trading frequency for the threshold $s_{0j}$ is
  \[
  \bar{f}_j = \frac{\sum_{t=1}^{T} \mathbb{1}\{z_t > s_{0j}\}}{T}.
  \]
- The empirical values $\bar{f}_j$ may not be a smoothed enough and the resulted profit function may not be accurate enough.
- Smooth the trading frequency function by regularization:
  \[
  \min_{f} \sum_{j=1}^{J} (\bar{f}_j - f_j)^2 + \lambda \sum_{j=1}^{J-1} (f_j - f_{j+1})^2
  \]
Optimum threshold $s_0$: Non-parametric approach

- The problem can be rewritten as

$$\min_{\bar{f}} \|\bar{f} - f\|^2_2 + \lambda \|Df\|^2_2$$

where

$$D = \begin{bmatrix} 1 & -1 & \cdots & -1 \\ 1 & -1 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{(J-1) \times J}.$$  

- Setting the derivative of the objective w.r.t. $f$ to zero yields the optimal solution $f^* = (I + \lambda D^T D)^{-1} \bar{f}$.  

- The optimal threshold is the one maximizes the total profit:

$$s_0^* = \arg \max_{s_{0j} \in \{s_{01}, s_{02}, \ldots, s_{0J}\}} \{s_{0j} f_j\}.$$
Optimum threshold $s_0$: Non-parametric approach

- Optimal threshold $s_0^*$ maximizes the total profit:

(a) Probability of trades

(b) Profit of each trade

(c) Total profit

Theoretical
NonParam: empirical
NonParam: regularized
If the spread $z_t$ is stationary, it can be written as

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

where

- $\mu$ represents the equilibrium value and
- $\epsilon_t$ is a zero-mean residual.

Equivalently, it can be written as

$$y_{1t} = \mu + \gamma y_{2t} + \epsilon_t$$

which now has the typical form for linear regression.

Least squares (LS) regression over $T$ observations:

$$\minimize_{\mu, \gamma} \sum_{t=1}^{T} (y_{1t} - (\mu + \gamma y_{2t}))^2$$

By stacking the $T$ observations in the vectors $y_1$ and $y_2$, we can finally write:

$$\minimize_{\mu, \gamma} \|y_1 - (\mu 1 + \gamma y_2)\|^2$$
Using the estimated parameters $\hat{\mu}$ and $\hat{\gamma}$, we can compute the residuals $\hat{\epsilon}_t = y_{1t} - \hat{\mu} - \hat{\gamma}y_{2t}$.

Then, one has to decide whether the cointegration is acceptable or not so move to the trading part.

There are many well-defined mathematical tests for the stationarity of $\hat{\epsilon}_t$, e.g., augmented Dicky-Fuller (ADF) test, Johansen test, etc.

Total profit:

profit of each trade $\times$ number of trades

- profit of each trade is $s_0$
- number of trades is related to the zero crossings, which can be analyzed theoretically as well as empirically.

Ideally, we want residuals with large amplitude (variance) as well as a strong mean reversion because they directly affect the profit.
LS regression for pairs trading

- One good case: 😊

Z-score and trading signal for EWH vs EWZ

Cum P&L

D. Palomar (HKUST)
LS regression for pairs trading

- But also a bad case: 😞

Z-score and trading signal for EWY vs EWT

Cum P&L

D. Palomar (HKUST)
LS regression for pairs trading

- The problem with the LS regression is that it assumes that $\mu$ and $\gamma$ are constant.
- In practice, they can change with time, resulting in a spread that drifts from equilibrium never to revert back with huge potential losses.
- Thus, in practice, $\mu$ and $\gamma$ are time-varying and have to be tracked.
- How to track time-varying parameters?

- Of course... Kalman!!!
- Well, you can also try a rolling regression or exponential smoothing, but Kalman works better.
Recall the previous static relationship for cointegrated series $y_{1t}$ and $y_{2t}$:

$$y_{1t} = \mu + \gamma y_{2t} + \epsilon_t$$

Let’s make it time-varying:

$$y_{1t} = \mu_t + \gamma_t y_{2t} + \epsilon_t$$

Let’s further assume that the parameters $\mu_t$ and $\gamma_t$ change slowly over time:

$$\mu_{t+1} = \mu_t + \eta_{1t}$$
$$\gamma_{t+1} = \gamma_t + \eta_{2t}$$

Obviously, this fits nicely the Kalman framework!
Interlude: The Kalman filter

- Kalman filter consist of two equations that model the time-varying hidden state $x_t$ and the observations $y_t$:

$$x_{t+1} = T_t x_t + \eta_t$$
$$y_t = Z_t x_t + \epsilon_t$$

- The observation equation $y_t = Z_t x_t + \epsilon_t$ relates the observation $y_t$ to the hidden state $x_t$ as a linear relationship, where $Z_t$ is the time-varying observation matrix and $\epsilon_t$ is a zero-mean Gaussian error $\epsilon_t \sim \mathcal{N}(0, R)$ with covariance matrix $R$.

- The state transition equation $x_{t+1} = T_t x_t + \eta_t$ expresses the transition of the hidden state from $x_t$ to $x_{t+1}$ as a linear relationship, where $T_t$ is the time-varying transition matrix and $\eta_t$ is a zero-mean Gaussian error $\eta_t \sim \mathcal{N}(0, Q)$ with covariance matrix $Q$.

- The Kalman filter is extremely versatile in modeling a variety of real-life processes.\(^9\)

---

Kalman for pairs trading

- Kalman filter (state transition equation and observation equation):

\[ x_{t+1} = Tx_t + \eta_t \]
\[ y_{1t} = Z_t x_t + \epsilon_t \]

where

- \( x_t \triangleq \begin{bmatrix} \mu_t \\ \gamma_t \end{bmatrix} \) is the hidden state
- \( T \triangleq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \) is the state transition matrix
- \( \eta_t \sim \mathcal{N}(0, Q) \) is the i.i.d. state transition noise with \( Q = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \)
- \( Z_t \triangleq \begin{bmatrix} 1 & y_{2t} \end{bmatrix} \) is the observation coefficient matrix
- \( \epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2) \) is the i.i.d. observation noise
- Note that this is a time-varying Kalman filter since \( Z_t \) is time-varying.

- Parameters \( \sigma_1^2, \sigma_2^2, \sigma_\epsilon^2 \) can be estimated using the EM algorithm using historical data for calibration.

D. Palomar (HKUST)
Kalman for pairs trading

- Log-prices of ETFs EWH and EWZ:

![Log-prices of ETFs EWH and EWZ](image)
Kalman for pairs trading

- Tracking of $\mu$ and $\gamma$ by LS, rolling LS, and Kalman:

![Tracking of $\mu$](image1)

![Tracking of $\gamma$](image2)

D. Palomar (HKUST)
Spreads achieved by LS, rolling LS, and Kalman:
**Kalman for pairs trading**

- **Trading of spread from LS:**

  ![Graph](image)

  **Z-score and trading on spread based on LS**
  
  2000–08–01 / 2003–12–31

  ![Graph](image)

  **Cum P&L for spread based on LS**
  
  2000–08–01 / 2003–12–31

  D. Palomar (HKUST)
Kalman for pairs trading

- Trading of spread from rolling LS:

  ![Graph showing Z-score and trading on spread based on rolling-LS](image)

  ![Graph showing Cum P&L for spread based on rolling-LS](image)

D. Palomar (HKUST)
Kalman for pairs trading

- Trading of spread from Kalman:

Z-score and trading on spread based on Kalman

Cum P&L for spread based on Kalman

D. Palomar (HKUST)
Kalman for pairs trading

- Wealth comparison:

![Cumulative P&L graph](image)

- LS
- rolling LS
- Kalman
The Kalman filter can and has been used in many aspects of financial time-series modeling as one could expect.\textsuperscript{10}

Examples of univariate time series: rate of inflation, national income, level of unemployment, etc.

Typical models include: local model, trend-cycle decompositions, seasonality, etc.

Examples of multivariate time series: inflation and national income.

Multiple time series allows for more sophisticated models including common factors, cointegration, etc.

Also data irregularities can be easily handled, e.g., missing observations, outliers, mixed frequencies.

Plenty of applications for nonlinear and non-Gaussian models as well, e.g., GARCH modeling and stochastic volatility modeling.

Outline

1. Cointegration
2. Basic Idea of Pairs Trading
3. Design of Pairs Trading
   - Pairs selection
   - Cointegration test
   - Optimum threshold
4. LS Regression and Kalman for Pairs Trading
5. From Pairs Trading to Statistical Arbitrage (StatArb)
   - VECM
   - Factor models
   - Optimization of mean-reverting portfolio (MRP)
6. Summary
Pairs trading focuses on finding cointegration between two stocks. A more general idea is to extend this statistical arbitrage from two stocks to more stocks. The idea is still based on cointegration: 

*Try to construct a linear combination of the log-prices of multiple (more than two) stocks such that it is a cointegrated mean-reversion process.*

In the case of two assets, the spread is \( z_t = y_{1t} - \gamma y_{2t} \), which can be understood as a portfolio with weights: \( \mathbf{w} = \begin{bmatrix} 1 \\ -\gamma \end{bmatrix} \).

In the general case of many assets, one has to properly design the portfolio \( \mathbf{w} \).
Outline

1. Cointegration
2. Basic Idea of Pairs Trading
3. Design of Pairs Trading
   - Pairs selection
   - Cointegration test
   - Optimum threshold
4. LS Regression and Kalman for Pairs Trading
5. From Pairs Trading to Statistical Arbitrage (StatArb)
   - VECM
   - Factor models
   - Optimization of mean-reverting portfolio (MRP)
6. Summary
Denote the log-prices of multiple stocks as $y_t$ and the log-returns as $r_t = \Delta y_t = y_t - y_{t-1}$.

Most of the multivariate time-series models attempt to model the log-returns $r_t$ (because the log-prices are nonstationary whereas the log-returns are weakly stationary, at least over some time horizon).

However, it turns out that differencing the log-prices may destroy part of the structure.

The VECM\textsuperscript{11} tries to fix that issue by including an additional term in the model:

$$r_t = \phi_0 + \Pi y_{t-1} + \sum_{i=1}^{p-1} \tilde{\Phi}_i r_{t-i} + w_t,$$

where the term $\Pi y_{t-1}$ is called error correction term.

The matrix $\Pi$ is of extreme importance.

Notice that from the model $r_t = \phi_0 + \Pi y_{t-1} + \sum_{i=1}^{p-1} \tilde{\Phi}_i r_{t-i} + w_t$ one can conclude that $\Pi y_t$ must be stationary even though $y_t$ is not!!!

If that happens, it is said that $y_t$ is cointegrated.

There are three possibilities for $\Pi$:

- $\text{rank}(\Pi) = 0$: This implies $\Pi = 0$, thus $y_t$ is not cointegrated (so no mystery here) and the VECM reduces to a VAR model on the log-returns.
- $\text{rank}(\Pi) = N$: This implies $\Pi$ is invertible and thus $y_t$ must be stationary already.
- $0 < \text{rank}(\Pi) < N$: This is the interesting case and $\Pi$ can be decomposed as $\Pi = \alpha \beta^T$ with $\alpha, \beta \in \mathbb{R}^{N \times r}$ with full column rank. This means that $y_t$ has $r$ linearly independent cointegrated components, i.e., $\beta^T y_t$, each of which can be used for pairs trading.
Suppose the stock $i$ is cointegrated with some tradable factors:

$$y_{it} = \pi_i^T y^f_t + w_{it}$$

where

- $y_{it}$ is the log-price of the stock $i$,
- $y^f_t$ is the log-price of the tradable factors,
- $\pi_i$ is the vector of loading coefficients
- $w_{it}$ is a stationary mean-reversion process.

It can also be written in a factor model form:

$$r_{it} = \pi_i^T f_t + \varepsilon_{it}$$

where

- $r_{it} = y_{it} - y_{i,t-1}$ is the log-return of stock $i$,
- $f_t = y^f_t - y^f_{t-1}$ is the log-returns of the tradable factors, and
- $\varepsilon_{it} = w_{it} - w_{i,t-1}$ is the specific noise.
Statistical arbitrage based on factor models

- Recall the factor model form expression

\[ r_{it} = \pi_i^T f_t + \varepsilon_{it} \]

- The idea now is to first properly select some tradable factors \( f_t \) and then test whether the cumulative summation of the resulted specific noise \( \varepsilon_{it} \), i.e., \( w_{it} = \sum_{j=0}^{t} \varepsilon_{ij} \), is stationary or not.

- If positive, then one can define a spread to be

\[
\begin{align*}
    z_{it} &= w_{it} = \sum_{j=0}^{t} (r_{ij} - \pi_i^T f_j) = \left[ 1 - \pi_i^T \right] \left( \sum_{j=0}^{t} \begin{bmatrix} r_{ij} \\ f_j \end{bmatrix} \right) \\
    &= \begin{bmatrix} 1 & -\pi_i^T \end{bmatrix} \begin{bmatrix} y_{it} \\ y_{tf} \end{bmatrix}
\end{align*}
\]
Some tradable examples\textsuperscript{12} of $f_t$ are the log-returns of

- (explicit factors) the sector ETFs and/or
- (hidden factors) several largest eigen-portfolios\textsuperscript{13}

Again, for each constructed cointegration component, one can study the spread and find the optimal trading thresholds as before.


\textsuperscript{13}A eigen-portfolio is a portfolio whose weight is an eigenvector of the covariance matrix of the stock returns.
Outline

1 Cointegration

2 Basic Idea of Pairs Trading

3 Design of Pairs Trading
   - Pairs selection
   - Cointegration test
   - Optimum threshold

4 LS Regression and Kalman for Pairs Trading

5 From Pairs Trading to Statistical Arbitrage (StatArb)
   - VECM
   - Factor models
   - Optimization of mean-reverting portfolio (MRP)

6 Summary
Mean-reverting portfolio (MRP)

- In the case of two assets, the spread is $z_t = y_{1t} - \gamma y_{2t}$, which can be understood as a portfolio with weights: $\mathbf{w} = \begin{bmatrix} 1 \\ -\gamma \end{bmatrix}$.

- In the general case of many assets, one has to properly design the portfolio $\mathbf{w}$.

- One interesting approach is based on a VECM modeling of the universe of stocks:
  - From the parameter $\beta$ contained in the low-rank matrix $\mathbf{\Pi} = \alpha \beta^T$ one can simply use any column of $\beta$ (even all of them).
  - Even better, $\beta$ defines a cointegration subspace and we can then optimize the portfolio within that cointegration subspace.\(^\text{14}\)

---

Consider the log-prices $y_t$ and use $\beta$ to extract several spreads $s_t = \beta^T y_t$.

Let’s now use a portfolio $w$ to extract the best mean-reverting spread from $s_t$ as $z_t = w^T s_t$.

To design the portfolio $w$ we have two main objectives (recall that total profit equals: profit of each trade $\times$ number of trades):

- we want large variance (profit of each trade): $w^T M_0 w$, where $M_i = \mathbb{E} \left[ (s_t - \mathbb{E}[s_t]) (s_{t+i} - \mathbb{E}[s_{t+i}])^T \right]$
- we want strong mean reversion (number of trades): many proxies exist like the Portmanteau statistics or crossing statistics.
For example, if we use the Portmanteau statistics as a proxy for the mean reversion, the problem formulation becomes:

\[
\begin{align*}
\text{minimize} & \quad w^T \sum_{i=1}^P \left( \frac{w^T M_i w}{w^T M_0 w} \right)^2 \\
\text{subject to} & \quad w^T M_0 w = \nu \\
& \quad w \in \mathcal{W}.
\end{align*}
\]

Using other proxies, the formulation can be expressed more generally as \(^{15}\)

\[
\begin{align*}
\text{minimize} & \quad w^T H w + \lambda \sum_{i=1}^P \left( w^T M_i w \right)^2 \\
\text{subject to} & \quad w^T M_0 w = \nu \\
& \quad w \in \mathcal{W}.
\end{align*}
\]

MRP in practice

- Observe several stock log-prices and the spreads obtained from $\beta$:
Observe several stock log-prices and the spreads obtained from $\beta$:
Outline

1. Cointegration
2. Basic Idea of Pairs Trading
3. Design of Pairs Trading
   - Pairs selection
   - Cointegration test
   - Optimum threshold
4. LS Regression and Kalman for Pairs Trading
5. From Pairs Trading to Statistical Arbitrage (StatArb)
   - VECM
   - Factor models
   - Optimization of mean-reverting portfolio (MRP)
6. Summary
First of all, we have discovered the concept of cointegration.

We have learned the basic idea of pairs trading for cointegrated assets:

- searching for a cointegrated spread is the first step
- making sure that the chosen spread remains cointegrated is key (cointegrated tests)
- obtaining the cointegration ratio $\gamma$ and the entering and exiting thresholds are important details.

We have learned of the use of Kalman (initially developed for tracking missiles) filtering for improved pairs trading.

We have briefly explored the extension of pairs trading (for two stocks) to statistical arbitrage (for more than two stocks):

- VECM modeling is an important multivariate time-series modeling tool
- sophisticated portfolio designs on the cointegration subspace are possible.
Thanks

For more information visit:

https://www.danielppalomar.com