Introduction to Algorithmic Trading Strategies
Lecture 4

Pairs Trading by Cointegration

Haksun Li
haksun.li@numericalmethod.com
www.numericalmethod.com

Outline

- Distance method
- Cointegration
- Stationarity
- Dickey–Fuller tests

References

- Pairs Trading: A Cointegration Approach. Arlen David Schmidt. University of Sydney. Finance Honours Thesis. November 2008.
- Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Soren Johansen. Oxford University Press, USA. February 1, 1996.

Pairs Trading

- Definition: trade one asset (or basket) against another asset (or basket)
  - Long one and short the other
- Intuition: For two closely related assets, they tend to "move together" (common trend). We want to buy the cheap one and sell the expensive one.
  - Exploit short term deviation from long term equilibrium.
- Try to make money from "spread".

Spread

- $Z = X - \beta Y$
- β
  - hedge ratio
  - cointegration coefficient

Dollar Neutral Hedge

- Suppose ES (S&P500 E-mini future) is at 1220 and each point worth $50, its dollar value is about $61,000. Suppose NQ (Nasdaq 100 E-mini future) is at 1634 and each point worth $20, its dollar value is $32,680.
- $\beta = \frac{61000}{32680} = 1.87.$
- $Z = ES - 1.87 \times NQ$
- Buy Z = Buy 10 ES contracts and Sell 19 NQ contracts.
- Sell Z = Sell 10 ES contracts and Buy 19 NQ contracts.

Market Neutral Hedge

- Suppose ES has a beta of 1.25, NQ 1.11.
- We use $\beta = \frac{1.25}{1.11} = 1.13$

Dynamic Hedge

- β changes with time, covariance, market conditions, etc.
- Periodic recalibration.

Distance

- The distance between two time series:
  - $d = \sum (x_i - y_j)^2$
- x_i, y_j are the normalized prices.
- We choose a pair of stocks among a collection with the smallest distance, d.

Distance Trading Strategy

- Sell Z if Z is too expensive.
- Buy Z if Z is too cheap.
- How do we do the evaluation?

Z Transform

- We normalize Z.
- The normalized value is called z-score.
- $z = \frac{x-\bar{x}}{\sigma_x}$
- Other forms:
- $z = \frac{x-M\times\bar{x}}{S\times\sigma_x}$
  - M, S are proprietary functions for forecasting.

A Very Simple Distance Pairs Trading

- Sell Z when $z > 2$ (standard deviations).
  - Sell 10 ES contracts and Buy 19 NQ contracts.
- Buy Z when $z < -2$ (standard deviations).
  - Buy 10 ES contracts and Sell 19 NQ contracts.

Pros of the Distance Model

- Model free.
- No mis-specification.
- No mis-estimation.
- Distance measure intuitively captures the LOP idea.

Cons of the Distance Model

- Does not guarantee stationarity.
- Cannot predict the convergence time (expected holding period).
- Ignores the dynamic nature of the spread process, essentially treat the spread as i.i.d.
- Using more strict criterions works for equity. In fixed income trading, we don't have the luxury of throwing away many pairs.

Risks in Pairs Trading

- Long term equilibrium does not hold.
- Systematic market risk.
- Firm specific risk.
- Liquidity.

Stationarity

- These ad-hoc β calibration does not guarantee the single most important statistical property in trading: stationarity.
- Strong stationarity: the joint probability distribution of $\{x_t\}$ does not change over time.
- Weak stationarity: the first and second moments do not change over time.
  - Covariance stationarity

Cointegration

- Cointegration: select a linear combination of assets to construct an (approximately) stationary portfolio.
- A stationary stochastic process is mean-reverting.
- Long when the spread/portfolio/basket falls sufficiently below a long term equilibrium.
- Short when the spread/portfolio/basket rises sufficiently above a long term equilibrium.

Objective

- Given two I(1) price series, we want to find a linear combination such that:
  - $z_t = x_t - \beta y_t = \mu + \varepsilon_t$
- ε_t is I(0), a stationary residue.
- μ is the long term equilibrium.
- Long when $z_t < \mu - \Delta$.
- Sell when $z_t > \mu + \Delta$.

Stocks from the Same Industry

- Reduce market risk, esp., in bear market.
  - Stocks from the same industry are likely to be subject to the same systematic risk.
- Give some theoretical unpinning to the pairs trading.
  - Stocks from the same industry are likely to be driven by the same fundamental factors (common trends).

Cointegration Definition

- $X_t \sim \mathrm{CI}(d, b)$ if
  - All components of X_t are integrated of same order d.
  - There exists a β_t such that the linear combination, $\beta_t X_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_n X_{nt}$, is integrated of order $(d - b), b > 0$.
- β is the cointegrating vector, not unique.

Illustration for Trading

- Suppose we have two assets, both reasonably I(1), we want to find β such that
  - $Z = X + \beta Y$ is I(0), i.e., stationary.
- In this case, we have $d = 1, b = 1$.

A Simple VAR Example

- $y_t = a_{11}y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt}$
- $z_t = a_{21}y_{t-1} + a_{22}z_{t-1} + \varepsilon_{zt}$
- Theorem 4.2, Johansen, places certain restrictions on the coefficients for the VAR to be cointegrated.
  - The roots of the characteristics equation lie on or outside the unit disc.

Coefficient Restrictions

- $a_{11} = \frac{(1 - a_{22}) - a_{12}a_{21}}{1 - a_{22}}$
- $a_{22} > -1$
- $a_{12}a_{21} + a_{22} < 1$

VECM (1)

- Taking differences
  - $y_t - y_{t-1} = (a_{11} - 1)y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt}$
  - $z_t - z_{t-1} = a_{21}y_{t-1} + (a_{22} - 1)z_{t-1} + \varepsilon_{zt}$
- $\begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} a_{11} - 1 & a_{12} \\ a_{21} & a_{22} - 1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$
- Substitution of a_{11}
  - $\begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} \frac{-a_{12}a_{21}}{1-a_{22}} & a_{12} \\ a_{21} & a_{22} - 1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$

VECM (2)

- $\Delta y_t = \alpha_y(y_{t-1} - \beta z_{t-1}) + \epsilon_{yt}$
- $\Delta z_t = \alpha_z(y_{t-1} - \beta z_{t-1}) + \epsilon_{zt}$
- $\alpha_y = \frac{-a_{12}a_{21}}{1-a_{22}}$
- $\alpha_z = a_{21}$
- $\beta = \frac{1-a_{22}}{a_{21}}$, the cointegrating coefficient
- $y_{t-1} - \beta z_{t-1}$ is the long run equilibrium, I(0).
- α_y, α_z are the speed of adjustment parameters.

Interpretation

- Suppose the long run equilibrium is 0,
  - $\Delta y_t, \Delta z_t$ responds only to shocks.
- Suppose $\alpha_y < 0$, $\alpha_z > 0$,
  - $\{y_t\}$ decreases in response to a +ve deviation.
  - $\{z_t\}$ increases in response to a +ve deviation.

Granger Representation Theorem

- If X_t is cointegrated, an VECM form exists.
- The increments can be expressed as a functions of the dis-equilibrium, and the lagged increments.
- $\Delta X_t = \alpha\beta' X_{t-1} + \sum c_t \Delta X_{t-1} + \varepsilon_t$
- In our simple example, we have
  - $\begin{bmatrix}\Delta y_t \\ \Delta z_t\end{bmatrix} = \begin{bmatrix}\alpha_y \\ \alpha_z\end{bmatrix}\begin{bmatrix}1 & -\beta\end{bmatrix}\begin{bmatrix}y_{t-1} \\ z_{t-1}\end{bmatrix} + \begin{bmatrix}\varepsilon_{yt} \\ \varepsilon_{zt}\end{bmatrix}$

Granger Causality

- $\{z_t\}$ does not Granger Cause $\{y_t\}$if lagged values of $\{\Delta z_{t-i}\}$ do not enter the Δy_t equation.
- $\{y_t\}$ does not Granger Cause $\{z_t\}$if lagged values of $\{\Delta y_{t-i}\}$ do not enter the Δz_t equation.

Test for Stationarity

- An augmented Dickey–Fuller test (ADF) is a test for a unit root in a time series sample.
- It is an augmented version of the Dickey–Fuller test for a larger and more complicated set of time series models.
- Intuition:
  - if the series y_t is stationary, then it has a tendency to return to a constant mean. Therefore large values will tend to be followed by smaller values, and small values by larger values. Accordingly, the level of the series will be a significant predictor of next period's change, and will have a negative coefficient.
  - If, on the other hand, the series is integrated, then positive changes and negative changes will occur with probabilities that do not depend on the current level of the series.
  - In a random walk, where you are now does not affect which way you will go next.

ADF Math

- $\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p-1} \Delta y_{t-i} + \epsilon_t$
- Null hypothesis $H_0: \gamma = 0$. (y_t non-stationary)
- $\alpha = 0, \beta = 0$ models a random walk.
- $\beta = 0$ models a random walk with drift.
- Test statistics = $\frac{\hat{\gamma}}{\sigma(\hat{\gamma})}$, the more negative, the more reason to reject H_0 (hence y_t stationary).
- SuanShu: AugmentedDickeyFuller.java

Engle-Granger Two Step Approach

- Estimate either
  - $y_t = \beta_{10} + \beta_{11} z_t + e_{1t}$
  - $z_t = \beta_{20} + \beta_{21} y_t + e_{2t}$
  - As the sample size increase indefinitely, asymptotically a test for a unit root in $\{e_{1t}\}$ and $\{e_{2t}\}$ are equivalent, but not for small sample sizes.
- Test for unit root using ADF on either $\{e_{1t}\}$ and $\{e_{2t}\}$.
- If $\{y_t\}$ and $\{z_t\}$ are cointegrated, $\{\beta\}$ super converges.

Engle-Granger Pros and Cons

- Pros:
  - simple
- Cons:
  - This approach is subject to twice the estimation errors. Any errors introduced in the first step carry over to the second step.
  - Work only for two I(1) time series.

Testing for Cointegration

- Note that in the VECM, the rows in the coefficient, Π, are **NOT** linearly independent.
  - $\begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} \frac{-a_{12}a_{21}}{1-a_{22}} & a_{12} \\ a_{21} & a_{22}-1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$
- $\begin{bmatrix} \frac{-a_{12}a_{21}}{1-a_{22}} & a_{12} \end{bmatrix} \times \frac{-(1-a_{22})}{a_{12}} = \begin{bmatrix} a_{21} & a_{22}-1 \end{bmatrix}$
- The rank of Π determine whether the two assets $\{y_t\}$ and $\{z_t\}$ are cointegrated.

VAR & VECM

- In general, we can write convert a VAR to an VECM.
- VAR (from numerical estimation by, e.g., OLS):
  - $X_t = \sum_{i=1}^{p} A_i X_{t-i} + \varepsilon_t$
- Transitory form of VECM (reduced form)
  - $\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t$
- Long run form of VECM
  - $\Delta X_t = \sum_{i=1}^{p-1} \Upsilon_i \Delta X_{t-i} + \Pi X_{t-p} + \varepsilon_t$

The Π Matrix

- Rank(Π) = n, full rank
  - The system is already stationary; a standard VAR model in levels.
- Rank(Π) = 0
  - There exists NO cointegrating relations among the time series.
- 0 < Rank(Π) < n
  - $\Pi = \alpha\beta'$
  - β is the cointegrating vector
  - α is the speed of adjustment.

Rank Determination

- Determining the rank of Π is amount to determining the number of non-zero eigenvalues of Π.
  - Π is usually obtained from (numerical VAR) estimation.
  - Eigenvalues are computed using a numerical procedure.

Trace Statistics

- Suppose the eigenvalues of Π are: $\lambda_1 > \lambda_2 > \cdots > \lambda_n$.
- For the o eigenvalues, $\ln(1 - \lambda_i) = 0$.
- For the (big) non-zero eigenvalues, $\ln(1 - \lambda_i)$ is (very negative).
- The likelihood ratio test statistics
  - $Q\big(H(r)|H(n)\big) = -T\sum_{i=r+1}^{p} \log(1 - \lambda_i)$
  - Ho: rank ≤ r; there are at most r cointegrating β.

Test Procedure

- int r = 0;//rank
- for (; r <= n; ++r) {
  - compute Q = $Q\left(H(r)|H(n)\right)$;
  - If (Q > c.v.) {//compare against a critical value
    - break;//fail to reject the null hypothesis; rank found
  - }
- }
- r is the rank found

Decomposing Π

- Suppose the rank of $\Pi = r$.
- $\Pi = \alpha\beta'$.
- Π is $n \times n$.
- α is $n \times r$.
- β' is r $\times$ n.

Estimating β

- β can estimated by maximizing the log-likelihood function in Chapter 6, Johansen.
  - $\log L(\Psi, \alpha, \beta, \Omega)$
- Theorem 6.1, Johansen: β is found by solving the following eigenvalue problem:
  - $\left|\lambda S_{11} - S_{10} {S_{00}}^{-1} S_{01}\right| = 0$

β

- Each non-zero eigenvalue λ corresponds to a cointegrating vector, which is its eigenvector.
- $\beta = (v_1, v_2, \cdots, v_r)$
- β spans the cointegrating space.
- For two cointegrating asset, there are only one β (v_1) so it is unequivocal.
- When there are multiple β, we need to add economic restrictions to identify β.

Trading the Pairs

- Given a space of (liquid) assets, we compute the pairwise cointegrating relationships.
- For each pair, we validate stationarity by performing the ADF test.
- For the strongly mean-reverting pairs, we can design trading strategies around them.