Multidimensional Pair Trading in China's Stock Market

Feng Yueman
Department of Statistics and Actuarial Science, HKU

Research Colloquium for Science UG Students 2022-23

Name: Feng Yueman
University No.: 3035447130
Major: Decision Analytics
Supervisor: Dr. Chen Wang

Abstract

Pair trading is a classic approach to algorithmically trading financial assets such as equities and commodities. Traditional pair trading uses only 2 non-stationary time series to form cointegrated pair with stationarity, where mean-reversion method can be applied. This study extends the pair trading strategy to multi-dimensional cointegrated set of financial time series and test the effectiveness of cointegration on China's stock market, particularly, CSI 300 Index component stocks.

Introduction

Linear regression is a commonly used method to find the relationship between several time series processes. However, Granger and Newbold[1] argued that linear regression analysis might lead to spurious regression, a situation in which the regression relationship appears to be statistically significant when variables are unrelated. Many economic time series exhibit random walk and other non-stationary behavior, leading to a higher possibility of finding spurious correlation under regression analysis. On the contrast, cointegration theory provides an alternative approach to study non-stationary time series. When a time series is not stationary, there is no tendency for its level to return to a constant mean over time; moreover, the volatility of the process is expected to grow boundlessly in the long run and any prediction based on the historical observations is not possible. Cointegration theory allows the identification of integration among time series that have similar dynamics in the long run and the estimation of their relationships.

In this study, we will be using Johansen's approach to exam cointegration relationship among multiple time series of stock prices. For cointegrated stock groups, we expect the stock with least 7-day return rate to rebound and thus, we buy in and hold the stock until its bounce.

Methodology

Integration: y_t is $I(d)$ if its $(d-1)^{th}$ difference is $I(0)$. That is, $\Delta^d y_t$ is stationary.

Augmented Dickey-Fuller Test: ADF test is a common way to assess the integration of a time series. Assume time series y_t follows autoregressive process of order p ($AR(p)$):

$$\Delta y_t = C + \rho y_{t-1} + \sum_{i=2}^{p} \emptyset_i \Delta y_{t-(i-1)} + \varepsilon_t$$

where:

$$\begin{cases} \rho = \sum_{i=1}^{p} \alpha_i - 1 \\ \phi_i = -\sum_{j=i}^{p} \alpha_j \end{cases}$$

The null and alternative hypothesis are:

$$H_0\colon \rho = 0 \qquad H_1\colon \rho < 0$$

ADF test statistic is:

$$F_\tau = \frac{\hat{\rho}}{SE(\hat{\rho})}$$

Cointegration: An $m * 1$ vector time series Y_t is said to be cointegrated of order (d, b), $CI(d, b)$ where $0 < b \le d$, if each of its component series Y_{it} is I(d) but some linear combination $\alpha^T Y_t$ is $I(d-b)$ for some constant vector $\alpha \neq 0$.

Johansen's Cointegration Test: Estimate y_t by vector autoregressive (VAR) process with order p:

$$y_t = \Phi_1 y_{t-1} + \cdots \Phi_p y_{t-p} + a_t + u_t$$

The error correction model is:

$$\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + a_t + u_t$$

where:

$$\Pi = \sum_{i=1}^{p} \Phi_i - I \qquad \Gamma_i = -\sum_{j=i+1}^{p} \Phi_j$$

Let the rank of be r:

if $r = 0$: there is no cointegration relationships

if $0 < r < k$: there are r cointegration relationships

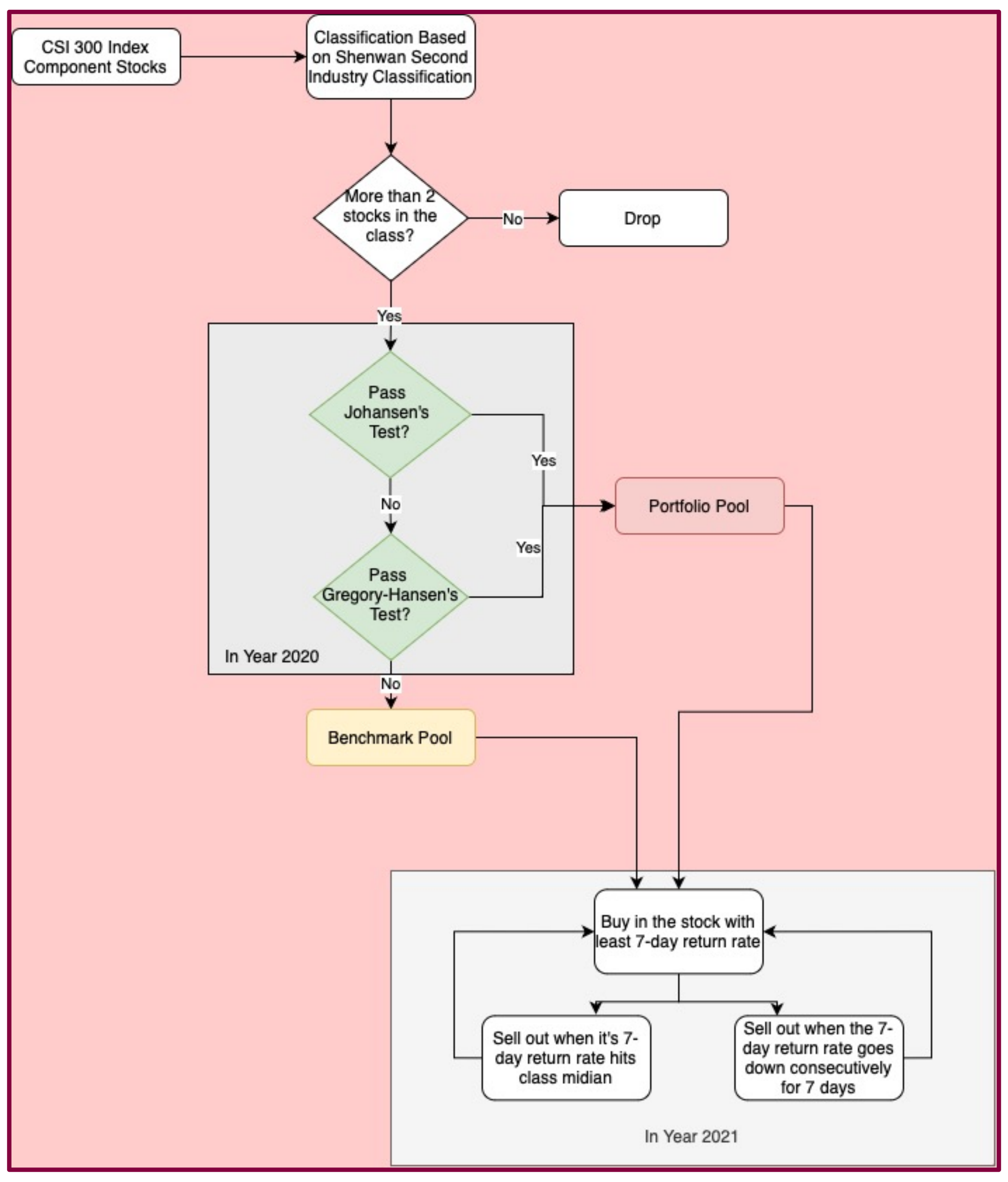
Gregory-Hansen Cointegration Test with Structural Breaks: structural breaks or regime shifts are common in financial time series due to market emergencies or major changes in policies and international relationships. Normal cointegration test assumes the cointegration to be both stationary in long-run mean and volatility. Gregory-Hansen's Cointegration Test allows one unknown structural breaks in:

(i) Level shift;

(ii) Level shift with trend;

(iii) Regime shift (both level and slope coefficient can change)

Results



Portfolio Pool: 300 stocks are categorized into 62 industrial classes and 37 of them have more than 2 stocks within each class. Among these 37 groups, 18 groups have passed either Johansen's Cointegration Test or Gregory-Hansen's Cointegration Test.

Johansen's Cointegration Test: the cointegration test result of industry group 'Insurance'

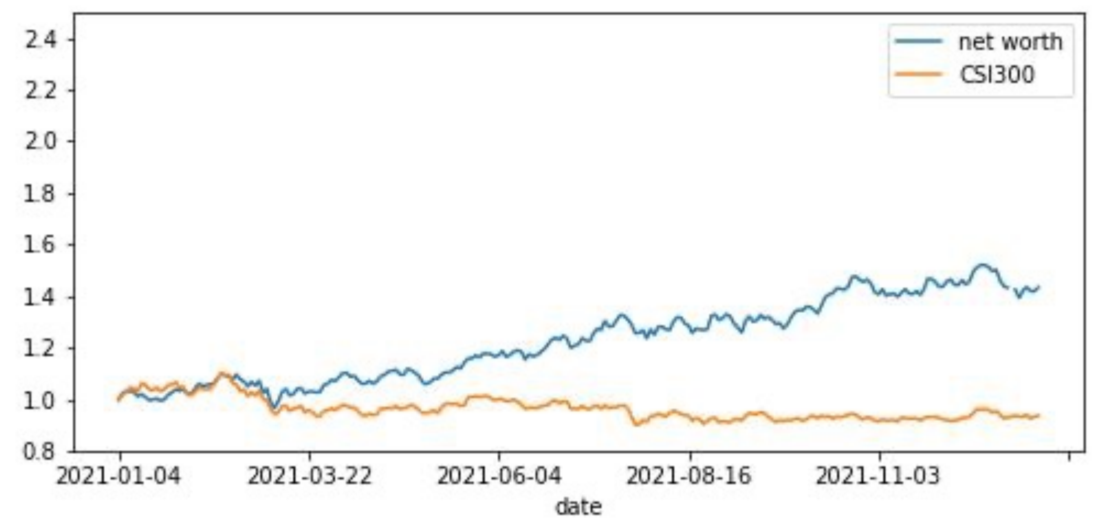
```
Johansen tests for cointegration
Trend: Constant                                    Number of obs  = 240
Sample: 05jan2020 thru 31aug2020                   Number of lags =   3
---------------------------------------------------------------------------
                                                                  Critical
Maximum                                                  Trace       value
   rank  Params          LL   Eigenvalue             statistic          5%
      0      55  -922.80642            .              75.1468        68.52
      1      64  -901.83957      0.16031              33.2131*       47.21
      2      71  -894.90204      0.05617              19.3380        29.68
      3      76  -889.64947      0.04283               8.8329        15.41
      4      79  -886.74984      0.02387               3.0336         3.76
      5      80  -885.23304      0.01256
---------------------------------------------------------------------------
* selected rank
```

Gregory-Hansen's Cointegration Test: the test result of industry group 'Insurance'

```
Gregory-Hansen Test for Cointegration with Regime Shifts
Model: Change in Regime and Trend                 Number of obs   =       243
Lags  =  2  chosen by Akaike criterion            Maximum Lags    =         8

              Test       Breakpoint   Date        Asymptotic Critical Values
            Statistic                                 1%        5%       10%
---------------------------------------------------------------------------
  ADF        -6.93           115     25apr2020     -7.31     -6.84     -6.58
  Zt         -7.17           116     26apr2020     -7.31     -6.84     -6.58
  Za        -82.94           116     26apr2020   -100.69    -88.47    -82.30
```

Investment Result: The annual return of investment for year 2021 is 43.4%, compared to 11.27% of investment return for non-cointegrated group and -7% for CSI 300 Index.



Discussion

- The cointegration filtration exhibits outstanding performance for gaining excess return compared to benchmark (CSI 300 Index)
- **Overrejection of Johansen's Cointegration Test when dimensionality is high:** it has been pointed out that the Johansen's cointegration test tends to overreject the null hypothesis (no cointegration relationship) when either the cross-sectional dimension (N) or time duration dimension (T) is high. Our study result generally supports this, but further research is expected.

N	3	4	5	6	7	8	9	10
Equally Weighted Return (no cointegrated class)	16.69%	20.86%	17.92%	5.90%	-23.42%	5.61%	-	12.20%
Equally Weighted Return (cointegrated class)	-17.89%	35.87%	-15.21%	25.31%	50.84%	79.55%	4.63%	2.01%

References: [1] Granger, C. W., & Newbold, P. (1974). Spurious regressions in econometrics. Journal of econometrics, 2(2), 111-120.
[2] Caldeira, J., & Moura, G. V. (2013). Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy. Available at SSRN 2196391..
[3] Chiu, M. C., & Wong, H. Y. (2015). Dynamic cointegrated pairs trading: Mean–variance time-consistent strategies. Journal of Computational and Applied Mathematics, 290, 516-534.
[4] Johansen, S. (1988). Statistical analysis of cointegration vectors. Journal of economic dynamics and control, 12(2-3), 231-254.
[5] Bykhovskaya, A., & Gorin, V. (2022). Cointegration in large VARs. The Annals of Statistics, 50(3), 1593-1617.