# Cointegration Method of Statistical Pairs Trading Arbitrage

Ruming Liu(ruminglia@usc.edu)
Junyue Han(junyueha@usc.edu)
Junxian Li(lijunxia@usc.edu)

*University of Southern California, Mathematical Finance*
Econ-515 Time Series Analysis Research
Instructor Ratika Narag

This paper discusses the implementation of statistical arbitrage in stock pairs trading using cointegration, Autogressive Moving Average model, Generalized Autoregressive Conditionally Heteroscedastic model and Vector Autoregressive model.

## INTRODUCTION

Pairs trading was first introduced in the mid-1980s by a group of technical analyst researchers that were employed by Morgan Stanley, the multinational investment bank and financial services company. The pairs trade strategy uses statistical and technical analysis to seek out potential market-neutral profits.

For example, Pepsi (PEP) and Coca-Cola (KO) are different companies that create a similar product, soda pop. Historically, the two companies have shared similar dips and highs, depending on the soda pop market. If the price of Coca-Cola were to go up a significant amount while Pepsi stayed the same, a pairs trader would buy Pepsi stock and sell Coca-Cola stock, assuming that the two companies would later return to their historical balance point. If the price of Pepsi rose to close that gap in price, the trader would make money on the Pepsi stock, while if the price of Coca-Cola fell, they would make money on having shorted the Coca-Cola stock.

The mathematical idea behind pairs trading is finding high correlated stocks which should track each other. When their spreads diverge, long the lower one and short the higher. Absolutely, if the spreads converge, you will earn the profit. As we know stock prices are usually non-stationary time series, using this pairs trading strategy we find the spreads are stationary. The reason behind is that these high correlated stocks usually have cointegration properties. And applying cointegration method can make a non-stationary series to a stationary series which avoids spurious regression problem. There are many ways to choose pairs, an easiest way is finding high correlated stocks, all stocks in this paper are selected by this method.

## TWO STOCKS COINTEGRATION ANALYSIS

For simplicity, we talk the simplest situation at the beginning, assume there are only two stocks. We denote them as $y_{1,t}$ and $y_{2,t}$. They have cointegration relation if there exist a vector $A$ such that,

$$y_{1,t} \sim I(d) \tag{1}$$

$$y_{2,t} \sim I(d) \tag{2}$$

$$Z_t = A' y_t \sim I(0) \tag{3}$$

Equation (1) & (2) tell us two stocks are difference stationary with the same of lag order, but their linear combination portfolio $Z_t$ is a stationary time series. We will use spread to represent this kind of portfolio later. Take stock pairs Kroger (KR) and United Rentals (URI) as our example. By OLS regression and Augmented Dickey-Fuller (ADF) test, $Z_t$ is stationary if $A' = [1 \quad 0.08904]$.

Then we will talk about cointegration for the Vector Autoregressive Representation. Suppose that $y_t$ can be represented as a non-stationary p-th order vector autoregression,

$$\Phi(L)y_t = \alpha + \epsilon_t \tag{4}$$

By Granger Representation Theorem we can write down Error-Correction Representation of this system, where

$$\Delta y_t = \xi_1 y_{t-1} + \xi_2 y_{t-2} \dots + \xi_p y_{t-p+1} + \alpha + \xi_0 y_{t-1} + \epsilon_t \tag{5}$$

$$\xi_s = -\left[\Phi_{s+1} + \cdots + \Phi_p\right], s = 1,2, \dots p-1 \tag{6}$$

$$\xi_0 = -\Phi(1) \tag{7}$$

With the data of KR & URI from 2015 to 2018, we solve the VAR (1) system below:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0.3948 \\ 2.5918 \end{bmatrix} - \begin{bmatrix} 0.9895 & 0 \\ -0.06 & 0.9922 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \tag{8}$$

It is a special case where $p = 1$, we can derive its Error-Correction Representation as,

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} = \begin{bmatrix} 0.3948 \\ 2.5918 \end{bmatrix} - \begin{bmatrix} 0.0105 & 0 \\ 0.06 & 0.0078 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \tag{9}$$

another way to get this ECR form is running the VAR (1) between vector $\Delta y_t$ & $y_{t-1}$, it should be the same with theoretical representation in (9) except computational error.

Using the same techniques in a different stock pair Paccar (PCAR) & Allstate (ALL), denote them as $y_{3,t}$ & $y_{4,t}$.

$$\begin{bmatrix} y_{3,t} \\ y_{4,t} \end{bmatrix} = \begin{bmatrix} 0.5069 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.9812 & 0.0068 \\ 0 & 0.995 \end{bmatrix} \begin{bmatrix} y_{3,t-1} \\ y_{4,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{3,t} \\ \epsilon_{4,t} \end{bmatrix} \tag{10}$$

$$\begin{bmatrix} \Delta y_{3,t} \\ \Delta y_{4,t} \end{bmatrix} = \begin{bmatrix} 0.5069 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.0188 & -0.0068 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{3,t-1} \\ y_{4,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{3,t} \\ \epsilon_{4,t} \end{bmatrix} \tag{11}$$

**FOUR STOCKS INTEGRATION ANALYSIS**

We have shown two different pairs stocks integration analysis, $y_{1,t}$ & $y_{2,t}$ has cointegration relation, $y_{3,t}$ & $y_{4,t}$ has cointegration relation. A natural question is whether $y_{1,t}$ & $y_{4,t}$ has a similar cointegration relation. An easy way to do that is run the regression between $y_{1,t}$ & $y_{4,t}$ like what we did in last part. Or equivalent, we can put these 4 stocks in one vector, applying the same techniques to get the vector $A'$, VAR model and ECM. The only difference is the dimension becomes 4 instead of 2.

$$A'' = \begin{bmatrix} 1 & 0.08904 & 0 & 0 \\ 1 & 0 & 0.50848 & 0 \\ 0 & 0 & 1 & -0.40799 \end{bmatrix} \tag{12}$$

Where $A''$ is the matrix which makes their linear combination $A''y_t$ stationary. By matrix transformation, we can easily know that:

$$A' = \begin{bmatrix} 1 & 0 & 0 & 0.2083 \\ 0 & 1 & 0 & -2.3299 \\ 0 & 0 & 1 & -0.40799 \end{bmatrix} \tag{13}$$

is the triangular form of matrix $A''$. We call $A'$ a basis for the space of cointegrating vectors. Any linear combination based on this basis is stationary time series. We can easily check $A'$ in last part is actually a combination of this basis. VAR (1) and VECM are:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \end{bmatrix} = \begin{bmatrix} 1.4301 \\ 0 \\ 1.3103 \\ 1.8000 \end{bmatrix} + \begin{bmatrix} 0.9803 & 0.0025 & -0.0124 & -0.0063 \\ 0 & 0.9871 & 0 & 0 \\ 0 & 0 & 0.9768 & 0 \\ -0.0274 & 0.0035 & 0 & 0.9841 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \\ y_{4,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \epsilon_{4,t} \end{bmatrix} \tag{14}$$

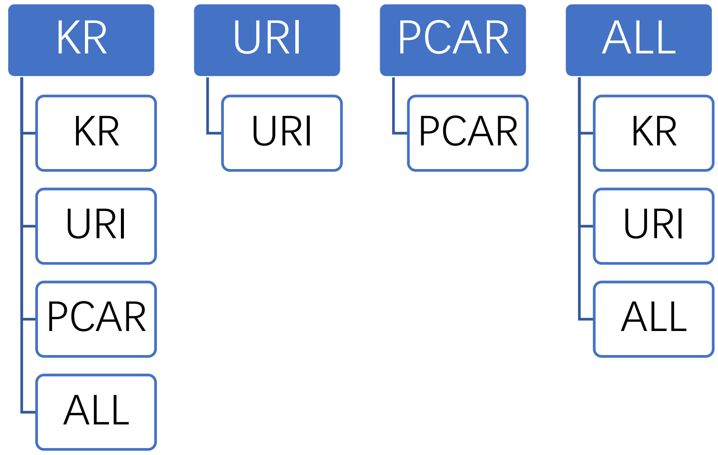
$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \\ \Delta y_{3,t} \\ \Delta y_{4,t} \end{bmatrix} = \begin{bmatrix} 1.4301 \\ 0 \\ 1.3103 \\ 1.8000 \end{bmatrix} - \begin{bmatrix} 0.0197 & -0.0025 & 0.0124 & 0.0063 \\ 0 & 0.0129 & 0 & 0 \\ 0 & 0 & 0.0232 & 0 \\ 0.0274 & -0.0035 & 0 & 0.0159 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \\ y_{4,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \epsilon_{4,t} \end{bmatrix} \tag{15}$$

With triangular basis in (13), we can write its Phillips's Triangular Representation too.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} -0.2083 \\ 2.3299 \\ 0.40799 \end{bmatrix} y_{4,t} + \begin{bmatrix} 1 & 0 & 0 & 0.2083 \\ 0 & 1 & 0 & -2.3299 \\ 0 & 0 & 1 & -0.40799 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \\ y_{4,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \epsilon_{4,t} \end{bmatrix} \tag{16}$$

## GRANGER CASUALITY TEST

As equation (14) we got, some of the coefficient are not significant different from 0, we can make a Granger casuality as below:



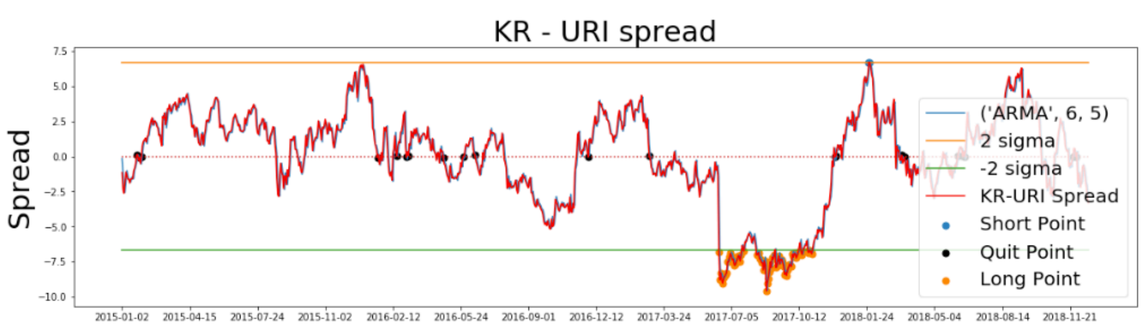

This Graph shows that the stock URI & PCAR are 'independent' with others, their stock prices are only determined by their past information. But stocks KR & ALL are different, their stock prices are not only determined by their past price information but also the past information of URI & PCAR, so we can say that URI & PCAR are some part reason which make KR & ALL changes.

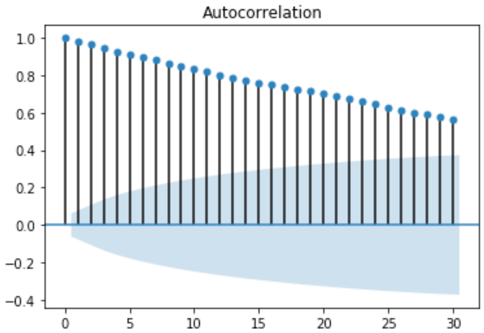
**AUTOGRESSIVE MOVING AVERAGE MODEL**

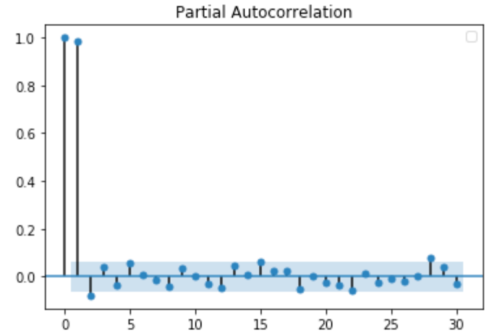
From the example of KR & URI cointegration analysis, we know if we set $A' = [1 \ \ 0.08904]$, then $A'y_t$ becomes a stationary time series which means $x_t \sim I(0)$ and

$$y_{1,t} + 0.08904y_{2,t} = 38.9733 + x_t \tag{17}$$

We also draw the graph of the past 3 years $x_t$ in equation (17), its ACF & PACF:






From the ACF, PACF and AIC criteria, we select ARMA (6, 5) as our regression model, and under 95% significance interval, our model of $x_t$ is,

$$x_t = -1.1359x_{t-2} + 1.7162\sigma_{t-2} + 1.3047\sigma_{t-3} + 0.8378\sigma_{t-4} + 0.7788\sigma_{t-5} + \sigma_t \quad (18)$$

ARMA Model Results

| | | | |
|---|---|---|---|
| Dep. Variable: | y | No. Observations: | 1000 |
| Model: | ARMA(6, 5) | Log Likelihood | -888.098 |
| Method: | css-mle | S.D. of innovations | 0.586 |
| Date: | Mon, 02 Nov 2020 | AIC | 1802.195 |
| Time: | 13:44:46 | BIC | 1865.996 |
| Sample: | 0 | HQIC | 1826.444 |

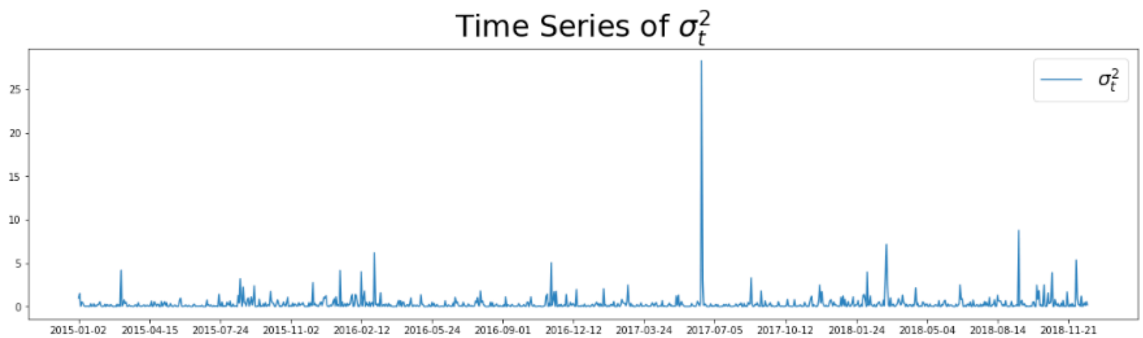
| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -0.1631 | 1.033 | -0.158 | 0.874 | -2.187 | 1.861 |
| ar.L1.y | 0.4249 | 0.375 | 1.134 | 0.257 | -0.309 | 1.159 |
| ar.L2.y | -1.1359 | 0.444 | -2.557 | 0.011 | -2.007 | -0.265 |
| ar.L3.y | 0.5188 | 0.762 | 0.681 | 0.496 | -0.975 | 2.013 |
| ar.L4.y | 0.2983 | 0.753 | 0.396 | 0.692 | -1.177 | 1.773 |
| ar.L5.y | 0.1177 | 0.442 | 0.266 | 0.790 | -0.749 | 0.985 |
| ar.L6.y | 0.6692 | 0.366 | 1.830 | 0.067 | -0.047 | 1.386 |
| ma.L1.y | 0.6348 | 0.336 | 1.889 | 0.059 | -0.024 | 1.294 |
| ma.L2.y | 1.7162 | 0.073 | 23.569 | 0.000 | 1.573 | 1.859 |
| ma.L3.y | 1.3047 | 0.624 | 2.090 | 0.037 | 0.081 | 2.528 |
| ma.L4.y | 0.8378 | 0.071 | 11.804 | 0.000 | 0.699 | 0.977 |
| ma.L5.y | 0.7788 | 0.335 | 2.326 | 0.020 | 0.122 | 1.435 |

Roots

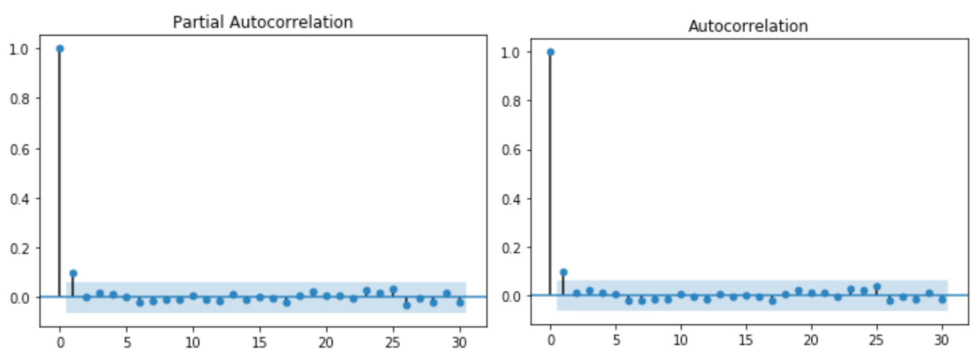
| | Real | Imaginary | Modulus | Frequency |
|---|---|---|---|---|
| AR.1 | 1.0186 | -0.0000j | 1.0186 | -0.0000 |
| AR.2 | 0.2418 | -0.9945j | 1.0235 | -0.2120 |
| AR.3 | 0.2418 | +0.9945j | 1.0235 | 0.2120 |
| AR.4 | -0.1491 | -0.9963j | 1.0074 | -0.2736 |
| AR.5 | -0.1491 | +0.9963j | 1.0074 | 0.2736 |
| AR.6 | -1.3799 | -0.0000j | 1.3799 | -0.5000 |
| MA.1 | -1.2516 | -0.0000j | 1.2516 | -0.5000 |
| MA.2 | 0.2299 | -0.9865j | 1.0129 | -0.2136 |
| MA.3 | 0.2299 | +0.9865j | 1.0129 | 0.2136 |
| MA.4 | -0.1420 | -0.9899j | 1.0000 | -0.2727 |
| MA.5 | -0.1420 | +0.9899j | 1.0000 | 0.2727 |

## GENERALIZED AUTOREGRESSIVE CONDITIONALLY HETEROSCEDASTIC

Next step is to check whether the $\sigma_t$ is White Noise. According ACF, PACF of $\sigma_t$ and Ljung-Box test, we conclude $\sigma_t$ is a white noise series. But it's not enough, we should continue check whether there exists ARCH effect in $\sigma_t$. At first, we plot the graph of $\sigma_t^2$,




And from the graph we can find there exists clustering problems, large $\sigma_t^2$ usually appears in same horizons. We also check the ACF & PACF of $\sigma_t^2$:




We decided to use GARCH(1, 1) model to estimate $\sigma_t$, GARCH summary is below:

```
                     Constant Mean - GARCH Model Results
==============================================================================
Dep. Variable:                   None   R-squared:                      -0.000
Mean Model:             Constant Mean   Adj. R-squared:                 -0.000
Vol Model:                      GARCH   Log-Likelihood:               -871.230
Distribution:                  Normal   AIC:                           1750.46
Method:            Maximum Likelihood   BIC:                           1770.09
                                        No. Observations:                 1000
Date:                Mon, Nov 02 2020   Df Residuals:                      996
Time:                        17:37:47   Df Model:                            4
                                Mean Model
=============================================================================
                 coef    std err          t      P>|t|       95.0% Conf. Int.
-----------------------------------------------------------------------------
mu         4.0157e-03  1.880e-02      0.214      0.831 [-3.284e-02,4.087e-02]
                             Volatility Model
===========================================================================
                 coef    std err          t      P>|t|     95.0% Conf. Int.
---------------------------------------------------------------------------
omega          0.1335  5.040e-02      2.650  8.058e-03 [3.475e-02,  0.232]
alpha[1]       0.1753      0.110      1.598      0.110 [-3.967e-02,  0.390]
beta[1]        0.4568      0.163      2.798  5.143e-03   [  0.137,  0.777]
===========================================================================
```

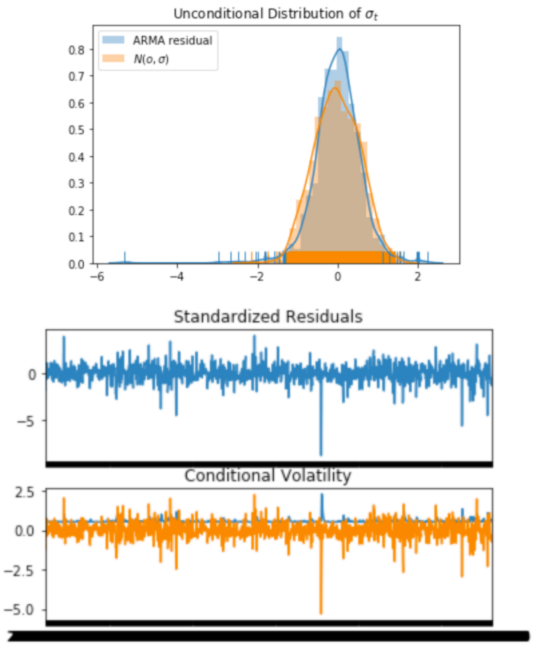
and with 90% confidence interval, our model is:

$$\sigma_t = \sqrt{h_t} v_t \tag{19}$$

$$v_t \sim N(0,1) \tag{20}$$

$$h_t = 0.1335 + 0.1753\sigma_{t-1}^2 + 0.4568h_{t-1}^2 \tag{21}$$

We also draw the PDF of $\sigma_t$ and compare it with samples from normal distribution with zero mean and same unconditional variance. Our PDF of $\sigma_t$ has a heavy tail and Kurtosis >3, these properties verify our GARCH model is reasonable to explain these phenomenons. In the second graph high conditional volatility $h_t$ appears with large $\sigma_t^2$ which is coordinate with the equation (21).




## TRADING STRATEGY MODEL

This part we will talk about our trading strategy of spreads. Like before, we use cointegration method to get the spread $x_t$ and model $x_t$ as an ARMA(6, 5)- GARCH(1, 1):

$$y_{1,t} + 0.08904 y_{2,t} = 38.9733 + x_t$$

$$x_t = -1.1359 x_{t-2} + 1.7162\sigma_{t-2} + 1.3047\sigma_{t-3} + 0.8378\sigma_{t-4} + 0.7788\sigma_{t-5} + \sigma_t$$

$$\sigma_t = \sqrt{h_t} v_t$$

$$h_t = 0.1335 + 0.1753\sigma_{t-1}^2 + 0.4568 h_{t-1}^2$$

How can we use the properties of this ARMA-GARCH model to make our trading decisions? We use AR(1) model to state the mean-reverting nature. If a time series $X_t \sim AR(1)$:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \eta_t \tag{22}$$

We can rewrite it as the following form:

$$X_t = X_{t-1} + \rho(\mu - X_{t-1}) + \eta_t \quad (23)$$

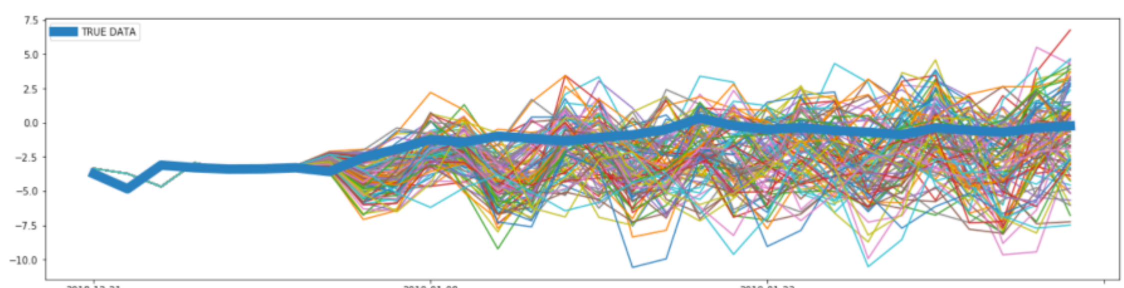
where $\mu = \frac{\phi_0}{1-\phi_1}$ and $\rho = (1 - \phi_1)$. This makes clear the mean-reverting nature of this process, and our ARMA-GARCH model also has a similar mean-reverting nature. This nature is the motivation of arbitrage in pairs trading. We use two investment strategies to make decision, one is long-term strategy, another is short-term strategy.

(1) Long-term strategy (Convergence Probability): Let $S_t$ denote the spread. We use $I_t$ represent the filtration generated by $\{S_t\}$. Then use Monte Carlo simulation to estimate $P[Convergence\ at\ terminal\ time|I_t]$. We enter a trade of the spread at time t if

$$P[Convergence\ at\ terminal\ time|I_t] > 0.5 \quad (24)$$

This simply means that we should trade only if the spread is likely to converge before the terminal time.

We use the first 30 trading day's data of 2019 as our forecast time window and run 1000 times Monte Carlo simulation to calculate its convergence probability at terminal time. Because our model is ARMA (6,5), we also need the last 6 prices data in 2018 as our initial price information $I_t$.




After calculation, our converge rate $P[Convergence\ at\ terminal\ time|I_t] = 0.79 > 0.5$. We decide to buy this spread because we think it has higher convergence probability than divergence probability. And the real situation is shown as the blue bold line, we can see that spread of KR & URI really converges at terminal time. This strategy works in the beginning of 2019.

(2) Short-term strategy (Growth Threshold): As stated earlier, a trade made at time t and closed at time t+1 will make profit if:

$$\begin{cases} S_t > S_{t+1}, & when\ S_t > 0 \\ S_t < S_{t+1}, & when\ S_t < 0 \end{cases} \tag{25}$$

Though it is impossible to know what will happen in time t+1, given the ARMA-GARCH model, we can compute the probability that it will converge at time t+1. If the probability is more than 0.8. We have high confidence that the spread will converge in next day. It is easy to see that given fitted model parameters,

$$S_t|I_t \sim \mathcal{N}(\phi_p + \sum_{i=1}^{6} \phi_i S_{t-i} + \sum_{i=1}^{5} \theta_i S_{t-i}\,, \mathrm{h}_t\,) \tag{26}$$

We will make a trade at time t if:

$$\begin{cases} P[S_t > S_{t+1}|I_t] > 0.8, when\ S_t > 0 \\ P[S_t < S_{t+1}|I_t] > 0.8, when\ S_t < 0 \end{cases} \tag{27}$$

Following the method we can make a decision table:

| Date | Original Data | Static Predict | error | ht | upper | lower | decision | real converge |
|---|---|---|---|---|---|---|---|---|
| 2018-12-21 | -3.75490406705822 | -3.360618497162453 | -0.394285569895767 | 0.0 | 0.0 | 0.0 | 0 | 0 |
| 2018-12-24 | -4.8459177074877005 | -3.733582738767058 | -1.1123349687206425 | 0.0 | 0.0 | 0.0 | 0 | 0 |
| 2018-12-26 | -3.1160102538992405 | -4.689120659784024 | 1.5731104058847833 | 0.0 | 0.0 | 0.0 | 0 | 0 |
| 2018-12-27 | -3.303998083822634 | -2.9501755061338604 | -0.3538225776887738 | 0.0 | 0.0 | 0.0 | 0 | 0 |
| 2018-12-28 | -3.40455061543139 | -3.4373969819058714 | 0.03284636647448158 | 0.0 | 0.0 | 0.0 | 0 | 0 |
| 2018-12-31 | -3.3898001557941058 | -3.1929553268980033 | -0.19684482889610244 | 0.0 | 0.0 | 0.0 | 0 | 0 |
| 2019-01-02 | -3.3070463580383347 | 3.9510365659346833 | 0.0714544848215466 | 0.14029250453204772 | 4.3255929762817065 | 3.57648015558766 | 1 | 1 |
| 2019-01-03 | -3.5830641636767595 | 4.260668106247324 | -0.3458020230884822 | 0.1433857684011882 | 4.6393312393896835 | 3.8820049731049644 | 1 | 0 |
| 2019-01-04 | -2.5591564630402814 | 1.7209426316149665 | 1.056973111702658 | 0.1638537753822105 | 2.1257311880023853 | 1.3161540752275478 | 1 | 1 |
| 2019-01-07 | -1.9537668354999482 | 4.227195381990702 | 0.37085871671460735 | 0.341607979122738 | 4.811667775797918 | 3.6427229881834857 | 1 | 1 |
| 2019-01-08 | -1.2437586234815825 | 5.8346457667789755 | 0.6974004556396336 | 0.21091683172255654 | 6.293902592229133 | 5.375388941328818 | 1 | 1 |
| 2019-01-09 | -1.4413777698652908 | 4.058100080108544 | -0.30207334437133815 | 0.23908136807986538 | 4.547059554993596 | 3.5691406052234926 | 1 | 0 |
| 2019-01-10 | -1.0001318633378986 | 1.8420156864308548 | 0.5062459422767229 | 0.17560647051020983 | 2.26106994366995 | 1.4229614291917594 | 1 | 1 |
| 2019-01-11 | -1.166660258064887 | 3.3018842974543703 | -0.29851749443750053 | 0.19251338296793205 | 3.7406477678663133 | 2.8631208270424273 | 1 | 0 |
| 2019-01-14 | -1.3549358149361197 | 5.5193614596326235 | -0.09213740300523288 | 0.16605110406088203 | 5.926855145547979 | 5.111867773717268 | 1 | 0 |
| 2019-01-15 | -1.0996398781872543 | 2.909973195770105 | 0.17387160220309106 | 0.14758350678321974 | 3.29413918435215 | 2.52580720718806 | 1 | 1 |
| 2019-01-16 | -0.9024727433229884 | 0.4007233510759587 | 0.17963209027429716 | 0.1487490640849539 | 0.7864033532603523 | 0.015043348891565056 | 1 | 1 |
| 2019-01-17 | -0.5138240632475259 | 3.0353534199958636 | 0.317447201039893 | 0.1492638122426289 | 3.421700171299183 | 2.6490066686925444 | 1 | 1 |
| 2019-01-18 | 0.3179881722315727 | 5.3728948646701715 | 0.8654918984868578 | 0.16134281917377394 | 5.774569881535515 | 4.971219847804828 | 1 | 1 |
| 2019-01-22 | -0.24016676086684186 | 3.3365555545441086 | -0.625024204766752 | 0.2767042540990777 | 3.862582411367102 | 2.8105286977211152 | 1 | 1 |
| 2019-01-23 | -0.514799392093849 | 1.6746953098103183 | -0.26820381716298597 | 0.2369568700394632 | 2.1614774685214435 | 1.1879131510991932 | 1 | 0 |
| 2019-01-24 | -0.3746273368016233 | 2.6027870641528166 | 0.1096283929488373 | 0.1717585667185667 | 3.0172247153364014 | 2.188349412969232 | 1 | 1 |
| 2019-01-25 | -0.5913514019816972 | 3.026547616098352 | -0.08891740842464302 | 0.1490828820041451 | 3.4126601411530895 | 2.640435091043615 | 1 | 0 |
| 2019-01-28 | -0.715093694889525 | 2.096655531694069 | -0.17636011517645023 | 0.1450386777246265 | 2.47749496994151 | 1.7158160934466278 | 1 | 0 |
| 2019-01-29 | -0.8916198576808938 | 2.0685862612450343 | -0.2631301609976895 | 0.14856168105534556 | 2.4540232613570443 | 1.6831492611330243 | 1 | 0 |
| 2019-01-30 | -0.39346090065246386 | 2.913149654281529 | 0.5157278101173897 | 0.15571916831118118 | 3.3077623356682038 | 2.5185369728948546 | 1 | 1 |
| 2019-01-31 | -0.5674276242612422 | 2.2041870307452047 | -0.13365383107653844 | 0.19120214426928972 | 2.641453704834701 | 1.7669203566557086 | 1 | 0 |
| 2019-02-01 | -0.7213726381840004 | 1.1400697994776732 | -0.17094336854436887 | 0.15333125780796572 | 1.5316451534109432 | 0.7484944455444031 | 1 | 0 |
| 2019-02-04 | -0.40169067069906106 | 5.757358698800033 | 0.23311727509425406 | 0.149362137466062 | 6.143832679122204 | 5.370884718477861 | 1 | 1 |
| 2019-02-05 | -0.2508060544257731 | 4.5460312338695115 | 0.18831521641630933 | 0.1532172174659052 | 4.93746094335246 | 4.154601524386563 | 1 | 1 |

At the beginning of each day, we use past data to calculate an confidence interval of today's spread, if yesterday spread is out of the interval, we will buy or sell this spread. In 'decision' column, we use 1 represents we open the position, 0 means we don't invest. In 'real converge' column, 1 means today's spread converges, 0 means it diverges. So if both columns are equal to 1, we will make profit from this trade. As the table shown, at the beginning of 2019, we make 20 pairs trade, 14 of them are profitable, our correct rate is 70% which is similar with our converge probability in last part. So this strategy also works in a short-term period investment.

**REFERENCES**


Krauss, Christopher. “STATISTICAL ARBITRAGE PAIRS TRADING STRATEGIES: REVIEW AND OUTLOOK.” *Journal of Economic Survey*, vol. 31, 2017, pp. 513–545.

Gatev, E., W. Goetzmann and K. Rouwenhorst Pairs Trading: Performance of a Relative-Value Arbitrage Rule, The Review of Financial Studies (2006), 19, 798- 827.

Stine, R. 2008 *Estimating an ARMA Process, lecture notes*, Time Series Analysis and Forecasting Stats 910, Wharton School of Finance

Qu, S., S. Mesmoeris, C. Davis et. al. ”Mean Reversion II: Pairs Trading Strategies”, Deutsche Bank Markets Research, (2016).

Ding, Z., C. W. J. Granger, and R. F. Engle. 1993. “A Long Memory Property of Stock Returns and a New Model.” *Journal of Empirical Finance* 1: 83–106.

Granger, C. W. J. 1969. “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods.” *Econometrica*, no. 37: 424–38.

Hamilton, James D. *Time Series Analysis*. Princeton University Press, 1994.

Jurek, Jakub, and Halla Yang. “Dynamic Portfolio Selection in Arbitrage.” *SSRN*, ssrn.com/abstract=882536.

Meucci, Attilio. “Review of Statistical Arbitrage, Cointegration, and Multivariate Ornstein-Uhlenbeck.” *SSRN*, 4 Dec. 2010, papers.ssrn.com/sol3/papers.cfm?abstract_id=1404905.