Efficient Pair Selection for Pair-Trading Strategies

Advanced Financial Data Analysis - Patrick McSharry

Assignment Module VII

Student Number 593233

April 4, 2015

Contents

Background

This research has been conducted with the support of my employer as a effort to automate the process of finding good stock pairs for quantitative strategies. As the utimate aim is to use this research in a production environment, all the coding work has been done with the internal language. This a proprietary language so the code is not reproduced here, however, pseudo-code is provided when appropriate.

Also great attention has been given to efficiency in testing for pairs, this is why we focused on the large-scale implementation of the Euler and Granger procedure.

The universe of stocks in consideration is the TOPIX 500 and we have used the last 7 years of daily prices data.

Introduction

Pair trading is a well-known and popular statistical arbitrage strategy. A pair is simply defined as two stocks that tend to move together (we need to define this notion more precisely). The strategy consists in trading the spread (a long position in one of the stocks vs a short position in the other) when a dislocation between the two prices paths is observed.

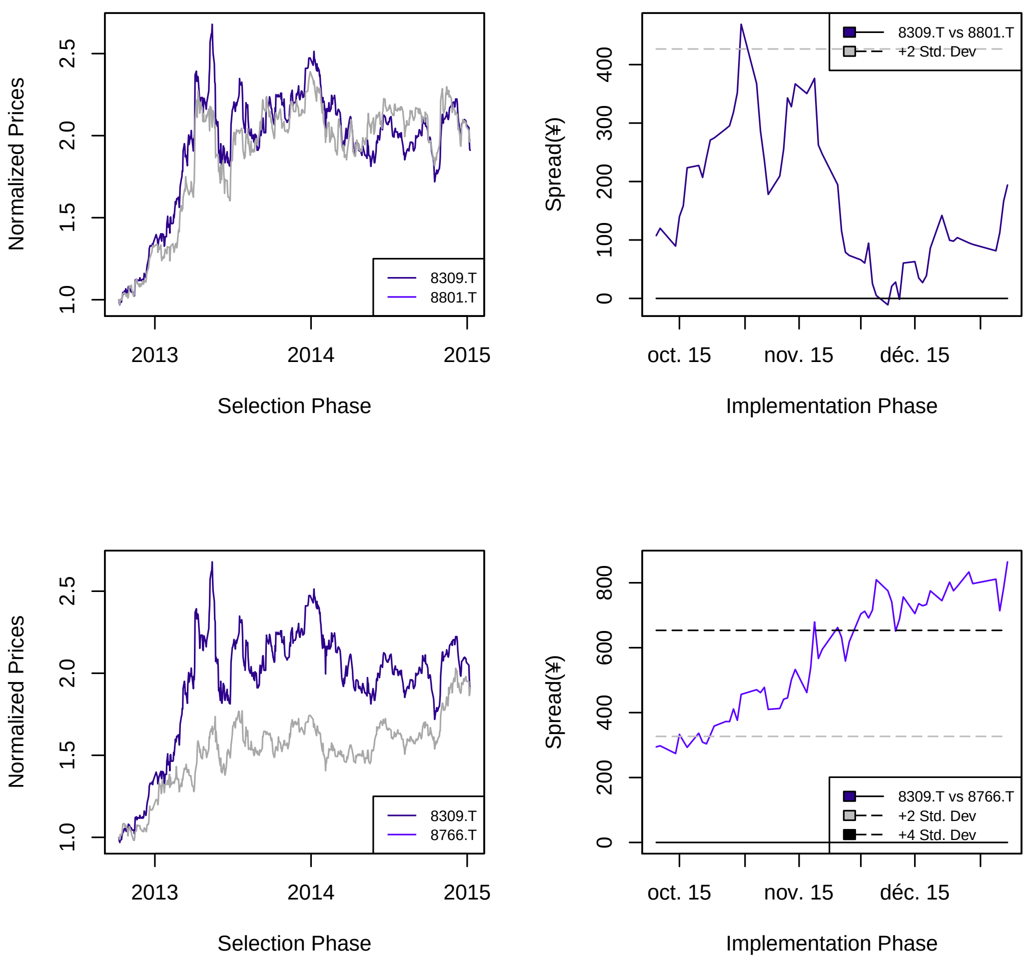
In the set-up of such a strategy, there are two distincts part: the selection part (what pairs we want to choose) and the implementation of the trading strategy (when/what sizes we shall trade). We here focus on the selection part. In order to backtest the selection process however, we need to have some simple trading rules to implement the strategy.

While those pairs can be chosen purely based on fundamental or statistical analysis, we will have a combined approach using both. The ultimate goal is to have an automated procedure to select for pairs that will generate positive PnL on average.

The most significant issues when chosing for pairs in a systematic pair-trading strategy are the testing procedure (how do we define assets that move together?) and the fact that the problem can get computationnally intense. There are about 50,000 different shares being traded in the world, the naive / brute-force way to handle the problem would require running $\binom{50000}{2} = 1,249,975,000$ tests!

In a first section, we formalize the concept of "assets moving together" and thereafter propose an efficient way to test a large number of pairs using the ADF Statistic. In a second section, we integrate some more fundamental factors to filter pair candidates before statitical testing, with the perspective of seeing what combination add the most value to the strategy. Lastly, the results will be presented and directions for improvement in the pair selecting model will be discussed.

Pair-trading - Two examples from the TOPIX 500



1 Testing multiple pairs with the Engle and Granger procedure

1.1 Correlation and cointegration

The first idea when one wants to test for assets that "moves together" is to use correlation of the returns. However, stongly-correlated returns do not signal much for long-term price paths, and at the same time the absence of correlation cannot be interpreted as independance. Also, from the implementation of a pair-trading strategy perspective, having a strong correlation may prevent from divergence and it reduces arbitrage opportunities. Correlation breaks may still be used for pair strategies, but pratitioners would only look at it for intraday / high-frequency trading. Our perspective is to set-up a medium term trading strategy (three month backtest period) and correlation will not help identify assets that "move together".

Engle and Granger [2] introduced the concept of cointegration.

A pair of assets X_t and Y_t is said to be cointegrated (of order 1) if:

- X_t and Y_t are integrated of order 1, i.e. they are $I(1)$
- There exists α, b such that the linear combination $Z_t = Y_t - \alpha X_t - b$ is $I(0)$ (i.e. the spread is stationary)

With the *a priori* knowledge that X_t and Y_t are $I(1)$, there are two steps in the Engle and Granger procedure:

1. Estimate the co-integrating relation (e.g. with OLS).
2. Test for stationarity of the residual (which we call spread here).

Cointegration enables to capture shared stochastic trends between several processes. With cointegrated assets, we can expect mean-reverting behaviour and by going long X_t / short Y_t when the spread is positive we should generate positive PnL once the spread gets back to its long-term level (naturally, we do the opposite trade when Z_t is negative).

As an aside, it should be noted that including the intercept term in the definition of cointegration should be considered carefully. We allowed for an intercept term to keep a general approach, however, this intercept term will not be traded traded once the pair strategy is set-up. In reality, we only take positions in X_t and Y_t. Also, in the implementation phase, there should some restriction on α. For example, $\alpha \leq 0$ implies that X_t and Y_t need to be bought or both sold, which does not make sense for a mean-reverting strategy. We can therefore disregard all the pairs with $\alpha \leq 0$.

1.2 ADF Test - Overview

There are quite a few methods to test for stationarity of the spread and we have implemented several of them (see Section 3 for further discussions).

The most popular test in the industry is the Augmented-Dickey-Fuller test. Below is the general set-up for the ADF test:

$$\Delta Z_t = \beta_0 + \beta_1 t + \delta Z_{t-1} + \sum_{i=1}^{k} \phi_i \Delta Z_{t-i} + \epsilon_t$$

BIC or AIC can be used to get the optimal lag k. After testing on some pairs, we have seen that including lags of order 2, 3, 4.... does not add much value while it significantly increases the complexity of the regression, hence the model with order 1 has been kept for all pairs.

In addition to the number of lag, the ADF Statistics varies depending on whether an intercept term / a linear trend are included. By construction, the spread should have zero mean and we could disregard the intercept term in the ADF formulation. However, we decided to follow what practitioners tend to do and kept that intercept term. Also, there could be a linear trend in the spread (for example if X_t grows faster than Y_t over the long term). In this case, we would need also need to include a linear trend in the ADF formulation. But this would require to have varying time-dependant hedge ratio in the implementation phase and this is not something we can afford (once we trade the pair, we do not want to amend the quantities every day).

Therefore, the ADF formulation we use from now on is:

$$\Delta Z_t = \beta + \delta Z_{t-1} + \phi \Delta Z_{t-1} + \epsilon_t$$

The null hypothesis of the ADF test is that the spread Z_t is a unit-root process, and the alternative is that the process is a stationary process. More formally, we test:

$$H_0 : \delta = 1 \qquad (X_t \text{ and } Y_t \text{ are not cointegrated})$$
$$H_1 : \delta < 1 \qquad (X_t \text{ and } Y_t \text{ are cointegrated})$$

The ADF Statistic is defined as the T-Statistic for the δ coefficient.

To get it for each pair we want to test, we need to:

1. Find the spread Z_t by OLS regression of Y_t on X_t
2. Find $\hat{\beta}, \hat{\delta}, \hat{\phi}$ by OLS regression of ΔZ_t on Z_{t-1} and ΔZ_{t-1}
3. Find $\hat{Var}(\hat{\delta})$
4. Compute the T-Stat for $\hat{\delta}$

In the Engle and Granger procedure, it should be noted that the usual ADF tabulated values cannot be used directly (as we test for stationarity on a derived variable - the residuals of an OLS). We use the values as estimated by MacKinnon [4] with 200 data points. All our tests use the 5% critical value.

Table 1: Critical values for Engle and Granger cointegration test

1pct	-3.95
5pct	-3.37
10pct	-3.07

1.3 Computation of the ADF Statistics

In order to avoid looping through the pairs and getting the ADF Statistics one by one, we need to rewrite the test in a matrix form. Below is the detailed procedure.

Assume we have M pairs, with each stock having N points of price history.

We write X^i and Y^i the column vectors for the price series of the pair i.

$$X = \begin{bmatrix} X_1^1 & X_1^2 & \dots & X_1^M \\ \vdots & & & \\ X_N^1 & X_N^2 & \dots & X_N^M \end{bmatrix} \quad Y = \begin{bmatrix} Y_1^1 & Y_1^2 & \dots & Y_1^M \\ \vdots & & & \\ Y_N^1 & Y_N^2 & \dots & Y_N^M \end{bmatrix}$$

1. Find the spread Z_t by OLS regression of Y_t on X_t

   We write:

$$\begin{bmatrix} Y_1^i \\ \vdots \\ Y_N^i \end{bmatrix} = \begin{bmatrix} X_1^i & 1 \\ \vdots & \vdots \\ X_N^i & 1 \end{bmatrix} \begin{bmatrix} \alpha^i \\ b^i \end{bmatrix} + \begin{bmatrix} u_1^i \\ \vdots \\ u_N^i \end{bmatrix}$$

   The OLS estimator for α^i and b^i is given by:

$$\begin{bmatrix} \hat{\alpha}^i \\ \hat{b}^i \end{bmatrix} = (\Gamma_1^{i'} \Gamma_1^i)^{-1} \Gamma_1^{i'} Y^i \text{ where } \Gamma_1^i = \begin{bmatrix} X_1^i & 1 \\ \vdots & \vdots \\ X_N^i & 1 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\alpha}^i \\ \hat{b}^i \end{bmatrix} = \frac{1}{N \sum (X_k^i)^2 - (\sum X_k^i)^2} \begin{bmatrix} N \sum X_k^i Y_k^i - \sum X_k^i \sum Y_k^i \\ - \sum X_k^i Y_k^i \sum X_k^i + \sum (X_k^i)^2 \sum Y_k^i \end{bmatrix}$$

(In the above, the sums are taken from $k = 0$ to $k = N$)

We thus get the spread:

$$Z = \begin{bmatrix} Z_1^1 & Z_1^2 & \dots & Z_1^M \\ \vdots & & & \\ Z_N^1 & Z_N^2 & \dots & Z_N^M \end{bmatrix} \text{ with } Z_k^i = Y_k^i - \hat{\alpha}^i X_k^i - \hat{b}^i$$

2. Find $\hat{\beta}, \hat{\delta}, \hat{\phi}$ by OLS regression of ΔZ_t on Z_{t-1} and ΔZ_{t-1}

We start by differentiating and lagging the spread and write the compact form for the regression. For the pair i, we get:

$$\begin{bmatrix} \Delta Z_3^i \\ \vdots \\ \Delta Z_N^i \end{bmatrix} = \begin{bmatrix} Z_2^i & \Delta Z_2^i & 1 \\ \vdots & \vdots & \vdots \\ Z_{N-1}^i & \Delta Z_{N-1}^i & 1 \end{bmatrix} \begin{bmatrix} \delta^i \\ \phi^i \\ \beta^i \end{bmatrix} + \begin{bmatrix} v_3^i \\ \vdots \\ v_N^i \end{bmatrix}$$

Again, the OLS estimator for δ^i, ϕ^i and β^i is given by:

$$\begin{bmatrix} \hat{\delta}^i \\ \hat{\phi}^i \\ \hat{\beta}^i \end{bmatrix} = (\Gamma_2^{i'} \Gamma_2^i)^{-1} \Gamma_2^{i'} \Delta Z^i \text{ where } \Gamma_2^i = \begin{bmatrix} Z_2^i & \Delta Z_2^i & 1 \\ \vdots & \vdots & \vdots \\ Z_{N-1}^i & \Delta Z_{N-1}^i & 1 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\delta}^i \\ \hat{\phi}^i \\ \hat{\beta}^i \end{bmatrix} = \frac{1}{Det} \begin{bmatrix} A_1 + A_2 + A_3 \\ A_4 + A_5 + A_6 \\ A_7 + A_8 + A_9 \end{bmatrix}$$

where:

$$\begin{aligned}
A_1 &= ((N-2)\sum(\Delta Z_{k-1}^i)^2 - (\sum \Delta Z_{k-1}^i)^2) \sum Z_{k-1}^i \sum \Delta Z_k^i \\
A_2 &= (-(N-2)\sum \Delta Z_{k-1}^i Z_{k-1}^i + \sum \Delta Z_{k-1}^i \sum Z_{k-1}^i) \sum \Delta Z_k^i \sum \Delta Z_{k-1}^i \\
A_3 &= (\sum \Delta Z_{k-1}^i Z_{k-1}^i \sum \Delta Z_{k-1}^i - \sum(\Delta Z_{k-1}^i)^2 \sum Z_{k-1}^i) \sum \Delta Z_k^i \\
A_4 &= (-(N-2)\sum \Delta Z_{k-1}^i Z_{k-1}^i + \sum \Delta Z_{k-1}^i \sum Z_{k-1}^i) \sum Z_{k-1}^i \sum \Delta Z_k^i \\
A_5 &= ((N-2)\sum(Z_{k-1}^i)^2 - \sum Z_{k-1}^i \sum Z_{k-1}^i) \sum \Delta Z_{k-1}^i \sum \Delta Z_k^i \\
A_6 &= (\sum \Delta Z_{k-1}^i Z_{k-1}^i \sum Z_{k-1}^i - \sum(Z_{k-1}^i)^2 \sum \Delta Z_{k-1}^i) \sum \Delta Z_k^i \\
A_7 &= (\sum \Delta Z_{k-1}^i Z_{k-1}^i \sum \Delta Z_{k-1}^i - \sum(\Delta Z_{k-1}^i)^2 \sum Z_{k-1}^i) \sum \Delta Z_k^i \sum Z_{k-1}^i \\
A_8 &= (\sum \Delta Z_{k-1}^i Z_{k-1}^i \sum Z_{k-1}^i - \sum(Z_{k-1}^i)^2 \sum \Delta Z_{k-1}^i) \sum \Delta Z_k^i \sum \Delta Z_{k-1}^i \\
A_9 &= (-(\sum \Delta Z_{k-1}^i Z_{k-1}^i)^2 + \sum(Z_{k-1}^i)^2 \sum(\Delta Z_{k-1}^i)^2) \sum \Delta Z_k^i
\end{aligned}$$

$$\begin{aligned}
Det^i = {} & (N-2)\Sigma(\Delta Z_{k-1}^i)^2 \Sigma(Z_{k-1}^i)^2 + \Sigma \Delta Z_{k-1}^i Z_{k-1}^i \Sigma \Delta Z_{k-1}^i \Sigma Z_{k-1}^i + \Sigma \Delta Z_{k-1}^i Z_{k-1}^i \Sigma \Delta Z_{k-1}^i \Sigma Z_{k-1}^i \\
& - \Sigma(\Delta Z_{k-1}^i)^2 (\Sigma Z_{k-1}^i)^2 - (N-2)(\Sigma \Delta Z_{k-1}^i Z_{k-1}^i)^2 - (N-2)(\Sigma \Delta Z_{k-1}^i Z_{k-1}^i)^2 - \Sigma(Z_{k-1}^i)^2 (\Sigma \Delta Z_{k-1}^i)^2
\end{aligned}$$

(Here the sums are taken from $k = 3$ to N or from $k = 2$ to $N - 1$ where appropriate).

3. Find $\hat{Var}(\hat{\delta})$

We have:

$$\hat{Var}(\begin{bmatrix}\hat{\delta}^i \\ \hat{\phi}^i \\ \hat{\beta}^i\end{bmatrix}) = (\hat{\sigma}^i)^2(\Gamma_2^{i'}\Gamma_2^i)^{-1} \text{ where } \hat{\sigma}^i \text{ is the MSE of the previous regression}$$

In this case, we only care about the standard error of the $\hat{\delta}^i$ coefficient.

$$\hat{Var}(\hat{\delta}^i) = \frac{(\hat{\sigma}^i)^2}{Det^i}((N-2)\sum(\Delta Z_{k-1}^i)^2 - (\sum \Delta Z_{k-1}^i)^2)$$

$$(\hat{\sigma}^i)^2 = \frac{1}{N-5}\sum(\Delta Z_k^i - \delta^i Z_{k-1}^i - \phi^i \Delta Z_{k-1}^i - \hat{\beta}^i)^2$$

In the above, we have $N-5$ appering as there are $N-2$ data points and the OLS removes 3 degrees of freedom.

4. Compute the T-Stat for $\hat{\delta}$

   The ADF Statistic for pair i is simply $\frac{\hat{\delta}^i}{\sqrt{\hat{Var}(\hat{\delta}^i)}}$

   A pair is considered to be cointegrated is the ADF Statistic is below the 5% critical value.

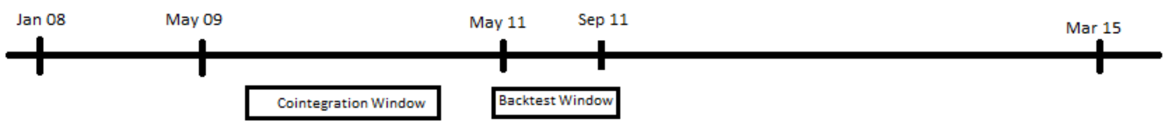
2 Filtering with fundamental factors and implementation

2.1 The pair selection and backtest procedure

As mentioned in the introductory section, a fairly long period of time has been used to set-up the pair selection procedure (last 7 years of data). We call the cointegration window the period over which we test for cointegration, and the backtest window the period over which we calculate the performance of the chosen pairs. Cointegration windows correspond to the estimation period, they are followed by the backtest windows (correspond to the trading period). The cointegration is estimated over a period of 2 years while the backtest length is 3 months. We take rolling windows for each (hence we have 20 cycles estimation / backtest - the estimation windows overlap).

Below is a simple scheme for one cointegration window / one backtest window and the pseudo-code for the selection and backtest procedure.

Cointegration and backtest windows



Pseudo-code for the selection and backtest procedure

```
For each cointegration/backtest period

  1. Filter the universe for pair candidates based on the factors

  2. Compute the ADF Statistic for all the pair candidates

  3. Remove all the pairs that do not pass the cointegration test and the hedge ratio test

  4. Run the backtest for those pairs

  5. Calculate performance and store the results

Next cointegration/backtest period
```

2.2 Filtering of pair candidates

Before any cointegration test, we need to bucket our universe. We will only look for cointegrated pairs inside each bucket. This step is required as cointegration tests (at least the ADF test) are not good at identifying spurious relationships (we will take a closer look at this problem in the section 3).

There should be an economic rationale behind each pair trade - i.e. we want the two legs of the pair to be connected due to fundamental reasons (having a similar business model is a good starting point).

This step is crucial and the ADF test should only be used to validate a relationship - it may not be the main discriminating tool.

We have tested several factors / factors combinations to bucket the TOPIX 500 universe:

- Sectors: stocks in the same sector tend to be affected by the same macro-economic factors. We have used the GICS® level 1 sector classification (10 sectors).
- Size: The market tends to be segmented by size: differents investor tend to focus on different size buckets. As a result, the long term price path of the small cap backet may be quite different from the one of the large cap bucket and we better keep pairs in the same bucket. The Size factor is proxied by the log of the market capitalisation.

- Value: The value factor is a well documented factor in the academic litterature to explain stock returns. A value stock can be seen as a "cheap" stock as its Price / Earnings ratio is below comparable stocks. On the other end, a stock with a relatively high P/E ratio can be seen as expensive.
  We can include the Value factor in our screening process ; however, in this case, we would rather chose stocks in different buckets than in the same bucket. More precisely, we need to go long the stock in the "cheap" bucket and short the stock in the "expensive" bucket so as to maximize the potential for mean-reversion.
  We used the Axioma Robust^{TM}Risk Model to proxy this factor.

- Volatility / Market Beta: Stocks with dissimilar volatility or beta profile are unlikely to obey to the same long term price trends. E.g. stocks with high beta / volatility reacts stronger in case of good / bad economic news. We may want to keep each leg of the pair in the same bucket.
  We used the Axioma Robust^{TM}Risk Model to proxy this factor (it could also be proxied by the variance of the log-returns / covariance of the log returns vs the market).

For the Size, Value and Volatility factors, 3 buckets are made. Stocks are classified according to those buckets.

This model can be extended to other factors, for which we may want to keep pairs in the same or in opposite buckets. Adding a factor should be justified by some economic intuition.

2.3 Backtesting methodology

Once the pairs candidates have been filtered and the cointegration tests run, there should be a backtesting phase. As mentioned previously, the backtesting window is 3 months and we have 20 of those in the overall backtest.
Backtesting the pairs requires to have some trading rules to decide when a position is opened and closed. This study focuses on the selection part so the trading rules are kept as simple as possible. The standard in the industry is to enter a position when the spread is ± 2 std. dev. away from the historical mean.

Below are the detailed trading rules:

1. Regress Y on X over the cointegration window to get the α and b. Compute the std. dev. σ of the spread Z
2. Calculate the hedge ratio as $H = \frac{\alpha X_{t_N}}{Y_{t_N}}$ with t_N the last day of the cointegration period. H represent the dollar amount of X that shall be bought or sold for every 1 dollar in Y
3. Each day t in the backtest window, compute the spread Z_t.
   - If $Z_t \geq 2\sigma$ go short Y and long X (at a ratio of H dollar of X for every dollar of Y)
   - If $Z_t \leq -2\sigma$ go long Y and short X
4. Close the position:
   - when Z_t is back to 0. This locks in a profit.
   - when $Z_t \geq 4\sigma$ or $Z_t \leq -4\sigma$. In this case, we have a loss

A basic risk management rule has been included, to avoid keeping the position open if prices continue to diverge. Also, we use the last day of the cointegration window to calculate the hedge ratio instead of the average over the whole window period in order to have pairs with net notional value close to \$0 (as the prices may have deviated quite a lot during the cointegration window). This constraint ensures we have a portfolio of pairs (fairly) market neutral.

In order to compare between different strategies, performance metrics are required.

Define the daily PnL for pair i as:

$$r_t^i = \log(\frac{Y_t^i}{Y_{t-1}^i}) - H\log(\frac{X_t^i}{X_{t-1}^i})$$

Assuming there are K days in the backtest window, a pair i is said to be winning if:

$$\sum_{k=1}^{K} r_{t_k}^i \geq 0$$

Also define the (daily) Sharpe Ratio for pair i as:

$$S^i = \frac{\bar{r^i}}{\sigma(r^i)} \text{ with } \bar{r^i} \text{ the average daily PnL and } \sigma(r^i) \text{ the volatility of the daily PnL}$$

Lastly, define the daily PnL for the strategy and (daily) Sharpe ratio for the strategy:

$$r_t = \frac{1}{M}\sum_{i=1}^{M} r_t^i \qquad\qquad S = \frac{\bar{r}}{\sigma(r)}$$

3 Results and further comments

3.1 Quicker cointegration testing

The major initial issue was the slowness to run multiple ADF tests. This has been overcome.

In the least-restricted pair candidates universe (with only a sector bucketting), there are around 18,000 pairs to test. Using a loop over the pre-packaged ADF function this would take 20 mins (similar time was required to run a batch with R).

The matrix implmentation of the ADF test that was described previously enabled to shrink the computation time by a factor of 40. It takes about 30 secs to test 18,000 pairs and output the ADF statistic, the α and b coefficients (we need to keep them as they are further required to compute the spread).

3.2 Performance analysis

We obtain equivocal results to the main question: what are the best factors to select for good pairs?

Below are the summary backtest results for all factor configurations, the detailed results for the bucketting by sector and the detailed results for the value factor. In the appendix can also be found the detailed results for the the combinations: sector and value, sector and size, sector and market beta.

Table 2: Summury for the entire backtest period and different factors configuration

Bucketting	**Nb Pairs**	**Nb Loosing**	**Nb Winning**	**Strategy Sharpe**
Sector	37,737	10,309	10,989	3.30%
Value	40,308	6,413	7,036	4.80%
Sector and Value	2,202	387	366	-0.85%
Sector and Size	13,005	3,540	3,775	2.45%
Sector and Beta	16,961	4,487	5,026	6.25%

Table 3: Bucketting by Sector

Start Backtest	**End Backtest**	**Nb Pairs**	**Nb Loosing**	**Nb Winning**	**Strategy Sharpe**
04-Jan-10	05-Apr-10	2,615	383	422	**5%**
02-Apr-10	02-Jul-10	2,578	392	526	**10%**
02-Jul-10	04-Oct-10	2,416	332	301	**1%**
04-Oct-10	04-Jan-11	2,513	293	516	**30%**
04-Jan-11	04-Apr-11	1,821	429	631	**17%**
05-Apr-11	05-Jul-11	1,458	231	391	**16%**
05-Jul-11	05-Oct-11	1,290	507	258	**-19%**
05-Oct-11	05-Jan-12	1,021	177	305	**15%**
05-Jan-12	05-Apr-12	1,266	329	332	**-1%**
05-Apr-12	05-Jul-12	1,378	378	399	**-1%**
05-Jul-12	05-Oct-12	1,135	408	314	**-18%**
05-Oct-12	07-Jan-13	1,090	358	338	**2%**
07-Jan-13	08-Apr-13	1,158	607	402	**-17%**
05-Apr-13	05-Jul-13	1,091	454	561	**16%**
05-Jul-13	07-Oct-13	2,505	760	1,056	**17%**
07-Oct-13	07-Jan-14	2,976	1,146	1,044	**-7%**
07-Jan-14	07-Apr-14	2,370	604	1,067	**31%**
08-Apr-14	08-Jul-14	2,995	901	874	**-2%**
08-Jul-14	08-Oct-14	2,417	929	713	**-18%**
08-Oct-14	08-Jan-15	1,644	691	539	**-11%**
Period Average					**3%**

Table 4: Bucketting by Value Factor

Start Backtest	End Backtest	Nb Pairs	Nb Loosing	Nb Winning	Strategy Sharpe
04-Jan-10	05-Apr-10	2,674	275	392	**12%**
02-Apr-10	02-Jul-10	2,676	261	259	**4%**
02-Jul-10	04-Oct-10	2,911	368	367	**1%**
04-Oct-10	04-Jan-11	2,507	68	317	**35%**
04-Jan-11	04-Apr-11	1,521	211	277	**2%**
05-Apr-11	05-Jul-11	1,305	99	349	**20%**
05-Jul-11	05-Oct-11	1,724	344	147	**-16%**
05-Oct-11	05-Jan-12	1,096	86	249	**21%**
05-Jan-12	05-Apr-12	1,329	146	163	**12%**
05-Apr-12	05-Jul-12	1,662	435	364	**-10%**
05-Jul-12	05-Oct-12	1,012	317	201	**-16%**
05-Oct-12	07-Jan-13	832	79	181	**33%**
07-Jan-13	08-Apr-13	1,119	214	169	**-21%**
05-Apr-13	05-Jul-13	1,098	323	338	**-1%**
05-Jul-13	07-Oct-13	2,635	440	380	**1%**
07-Oct-13	07-Jan-14	2,878	453	609	**18%**
07-Jan-14	07-Apr-14	2,147	406	572	**7%**
08-Apr-14	08-Jul-14	3,606	581	705	**3%**
08-Jul-14	08-Oct-14	3,224	812	453	**-16%**
08-Oct-14	08-Jan-15	2,352	495	544	**7%**
Period Average					**5%**

First, a few comments about the statistics that are displayed:

- "Nb pairs" corresponds to the number of pairs after candidates filtering (by factor) and cointegration test.
- "Start Backtest" and "End Backtest" correspond to the backtest window; the cointegration window occurs over the 2-year period before the "Start Backtest" date.
- In the summary table the "Strategy Sharpe" is the average Sharpe over the 20 backtest periods (this is a daily Sharpe).
- "Nb Pairs" $\neq$ "Nb Winning" +"Nb Loosing" as some of the pairs never breach the $\pm$ 2 std. dev. band and are not traded.
- Increasing the number of discriminating factors strongly reduces the number of pairs.

Now, some comments about the results themselves:

The performance results are overall positive: for most of the factor configurations, there are more winning pairs than loosing pairs and the Sharpe ratio is positive.

However, as it can be seen from the detailed results, the performance is quite volatile from one period to another. While positive, the Sharpe ratios are fairly small. For the best configuration (sector and beta bucketting) the daily Sharpe is 6.25%. This corresponds to an annualized Sharpe of 99% while 100%-150% is really the minimum Sharpe threshold to consider a quantitative strategy as interesting. (Trading costs would also need to be factored in, which can greatly reduce the profitability for a pair-trading strategy).

The results for the best factors are also not inline with expectations. From an economical point of view, the best configuration seemed to be the bucketting by sector and value. The pair trading with this bucketting can be summurized as "go long one cheap stock and short an expensive co-moving stock in the same sector". Unfortunately, the backtest shows that this strategy does not work well. On the other hand, the best performing strategy (bucketting by sector and beta) is not the most intuitive.

Even if those results do not show strong performance of the pair-trading strategies, we now have a methodology to test for other factor combinations and there should be some further interesting work on this side.

3.3 Cointegration and cointegration tests in question

Regardless of the implementation, the whole pair-trading strategy relies on cointegration to be stable. As for any strategy using historical data, we test over a past period and expect the model to remain valid in the next. We implicitly assume that pairs cointegrated in the past period will keep the same behaviour in the next period.

In the light of the generally dispointing results we got in the previous section, this assumption should be challenged and cointegration stability should be tested. A first indicator that something is going wrong can be seen in the huge variation of the number of pairs that pass the ADF tests (see Table 3, depending on the period, there may be from 1000 pairs to 3000 pairs with an ADF Statistics below the 5% significance value).

To investigate this further, we proceed to a simple test:

- Take a 4-year period, say Jan 2011 to Jan 2015 and divide it into two.
- Take a large number of pairs, says all the pairs given by bucketting by sector only - this gives 16,874 pairs.
- Check for cointegration separately in each one of the two periods.
- Calculate $P(C_1^{ADF} \cap C_2^{ADF})$ and $P(C_1^{ADF})P(C_2^{ADF})$ with C_k^{ADF} defined as "the pair is cointegrated during period k as per 5% significance in the ADF test".

For our sample period on the TOPIX 500, we get:

$$P(C_1^{ADF}) = \tfrac{4251}{16874} = 25.19\% \qquad P(C_2^{ADF}) = \tfrac{2310}{16874} = 13.69\% \qquad P(C_1^{ADF} \cap C_2^{ADF}) = \tfrac{579}{16874} = 3.43\%$$

As a result:

$$P(C_1^{ADF})P(C_2^{ADF}) = 3.45\% \simeq P(C_1^{ADF} \cap C_2^{ADF}) \qquad P(C_2^{ADF} \mid C_1^{ADF}) = 13.62\% \simeq P(C_2^{ADF})$$

C_1 and C_2 appear to be independant, and there is not such thing as a relationship $C_1 \Rightarrow C_2$.

This casts doubt on pair-trading overall, if no stationarity can be expected for the spread in the trading period, there is no reason for the mean-reverting strategy to generate positive performance.

Two things may be happening here: either there are structural breaks in the price paths and we cannot identify pairs that mean-revert or our test for cointegration is not appropriate.

As per Sjö [5] there are no test for unit-roots that is uniformly better than ADF. However, we still want to to check whether more stable cointegration measures can be obtained. The Phillips-Perron (PP) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests have been tried.

We briefly introduce their comparative advantages over ADF and run the previous simple test.

- The ADF deals with the autoregressive structure in the residuals (of the DF test) by introducing lagged differenced variable (s). We have assumed that including one lag was enough to obtain white-noise residuals. This may be challenged.

  Instead of including lags, the Phillips-Perron test uses a non-parametric correction so as the DF tabulated values for the T-Statistic remain correct. Alike ADF, the null hypothesis for the PP test is the presence of unit-root. In essence those tests are quite similar, but for the same data set and same level of significance, ADF rejects the null more frequently than PP (see below).

- When looking at cointegration, we actually want to check that the spread is a stationary process. For this purpose,it would make sense to use a test with stationarity as the null hypothesis. The KPSS test is designed in this way.

  It starts with the model:

$$
\begin{aligned}
z_t &= \Lambda' D_t + \mu_t + u_t \qquad & D_t \text{ contains deterministic components} \\
\mu_t &= \mu_{t-1} + v_t \qquad & \text{where } v_t \text{ is a white-noise of variance } \sigma^2
\end{aligned}
$$

  It tests the null of $\sigma^2 = 0$ vs $\sigma^2 < 0$.

Unfortunately those two tests do not obtain better results than the ADF test.

For the PP test:

$$P(C_1^{PP}) = 8.11\% \quad P(C_2^{PP}) = 4.78\% \quad P(C_1^{PP} \cap C_2^{PP}) = 0.48\% \quad P(C_1^{PP})P(C_2^{PP}) = 0.39\% \quad P(C_2^{PP} \mid C_1^{PP}) = 5.92\%$$

For the KPSS test:

$$P(C_1^{KPSS}) = 3.16\% \quad P(C_2^{KPSS}) = 2.67\% \quad P(C_1^{KPSS} \cap C_2^{KPSS}) = 0.03\%$$

$$P(C_1^{KPSS})P(C_2^{KPSS}) = 0.08\% \quad P(C_2^{KPSS} \mid C_1^{KPSS}) = 0.94\%$$

Again, those results show that cointegration for stock pairs is not persistent through time.

Conclusion

The main contribution of this study is the re-writing of the Engle and Granger procedure in a form that enables to test multiple pairs very rapidly. The procedure developped here also enabled to include easily some fundamental factors to guide the search for profitable pairs. In the optimized configuration (sector and market beta bucketting), the strategy generates a Sharpe ratio close to 1.

However, the tests that have been further made showed that cointegration may be an unreliable metric in a mean-reverting strategy. This leaves the door open to other ways to approach mean-reversion (in this perspective, we can think about modelling the spread with a Vector-Error-Correction-Model - VECM).

References

[1] Dickey, D. A. and Fuller, W. A *Distribution of the Estimators for Autoregressive Time Series with a Unit Root.* Journal of the American Statistical Association. 1979

[2] Engle, Robert and Granger, C.W.J. *Co-Integration and Error Correction: Representation, Estimation, and Testing.* Econometrica, Vol. 55. 1987.

[3] Wang, Jieren. *A high performance pair trading application.* Parallel & Distributed Processing. IEEE. 2009

[4] MacKinnon, J. G. *Critical values for cointegration tests.* Chapter 13 in Long-Run Economic Relationships: Readings in Cointegration. Oxford Oxford University Press. 1991.

[5] Sjö, Bo. *Testing for Unit Roots and Cointegration*, Aug 2008.

[6] Zivot, Eric. *Unit Root Tests.* Chapter 4, Lecture Notes on Time Series Econometrics. 2006.

Appendix

Table 5: Bucketting by Sector and Value Factor

Start Backtest	**End Backtest**	**Nb Pairs**	**Nb Loosing**	**Nb Winning**	**Strategy Sharpe**
04-Jan-10	05-Apr-10	169	22	29	**0%**
02-Apr-10	02-Jul-10	205	15	15	**-8%**
02-Jul-10	04-Oct-10	141	20	15	**4%**
04-Oct-10	04-Jan-11	125	2	22	**29%**
04-Jan-11	04-Apr-11	68	17	11	**-14%**
05-Apr-11	05-Jul-11	47	4	18	**20%**
05-Jul-11	05-Oct-11	61	17	5	**-19%**
05-Oct-11	05-Jan-12	55	3	11	**21%**
05-Jan-12	05-Apr-12	78	5	8	**11%**
05-Apr-12	05-Jul-12	81	19	14	**-6%**
05-Jul-12	05-Oct-12	48	13	8	**-10%**
05-Oct-12	07-Jan-13	29	4	9	**10%**
07-Jan-13	08-Apr-13	73	20	7	**-35%**
05-Apr-13	05-Jul-13	48	12	14	**6%**
05-Jul-13	07-Oct-13	147	22	15	**-9%**
07-Oct-13	07-Jan-14	221	45	40	**7%**
07-Jan-14	07-Apr-14	153	31	51	**9%**
08-Apr-14	08-Jul-14	212	47	33	**-8%**
08-Jul-14	08-Oct-14	140	47	20	**-21%**
08-Oct-14	08-Jan-15	101	22	21	**-4%**
Period Average					**-1%**

Table 6: Bucketting by Sector and Size

Start Backtest	End Backtest	Nb Pairs	Nb Loosing	Nb Winning	Strategy Sharpe
04-Jan-10	05-Apr-10	916	142	141	**-1%**
02-Apr-10	02-Jul-10	874	127	182	**12%**
02-Jul-10	04-Oct-10	871	120	112	**-2%**
04-Oct-10	04-Jan-11	857	94	174	**32%**
04-Jan-11	04-Apr-11	622	155	218	**15%**
05-Apr-11	05-Jul-11	491	68	140	**20%**
05-Jul-11	05-Oct-11	414	161	87	**-19%**
05-Oct-11	05-Jan-12	321	64	92	**11%**
05-Jan-12	05-Apr-12	458	114	117	**-2%**
05-Apr-12	05-Jul-12	477	113	144	**0%**
05-Jul-12	05-Oct-12	399	151	112	**-19%**
05-Oct-12	07-Jan-13	391	127	123	**5%**
07-Jan-13	08-Apr-13	399	215	140	**-17%**
05-Apr-13	05-Jul-13	333	134	177	**10%**
05-Jul-13	07-Oct-13	900	289	349	**11%**
07-Oct-13	07-Jan-14	1,044	399	367	**-6%**
07-Jan-14	07-Apr-14	850	222	374	**26%**
08-Apr-14	08-Jul-14	1,043	310	316	**1%**
08-Jul-14	08-Oct-14	802	308	233	**-18%**
08-Oct-14	08-Jan-15	543	227	177	**-10%**
Period Average					**2%**

Table 7: Bucketting by Sector and Market Beta

Sectors + Beta					
Start Backtest	**End Backtest**	**Nb Pairs**	**Nb Loosing**	**Nb Winning**	**Strategy Sharpe**
04-Jan-10	05-Apr-10	1,242	189	181	**3%**
02-Apr-10	02-Jul-10	1,219	191	243	**8%**
02-Jul-10	04-Oct-10	1,198	174	160	**1%**
04-Oct-10	04-Jan-11	1,138	139	253	**39%**
04-Jan-11	04-Apr-11	771	196	287	**16%**
05-Apr-11	05-Jul-11	543	76	134	**22%**
05-Jul-11	05-Oct-11	502	184	110	**-15%**
05-Oct-11	05-Jan-12	475	80	152	**19%**
05-Jan-12	05-Apr-12	677	171	191	**-1%**
05-Apr-12	05-Jul-12	746	199	222	**4%**
05-Jul-12	05-Oct-12	618	217	171	**-16%**
05-Oct-12	07-Jan-13	604	183	188	**7%**
07-Jan-13	08-Apr-13	626	319	223	**-13%**
05-Apr-13	05-Jul-13	466	187	238	**10%**
05-Jul-13	07-Oct-13	1,030	314	440	**18%**
07-Oct-13	07-Jan-14	1,126	430	421	**0%**
07-Jan-14	07-Apr-14	1,009	210	498	**41%**
08-Apr-14	08-Jul-14	1,250	386	379	**6%**
08-Jul-14	08-Oct-14	1,025	371	302	**-19%**
08-Oct-14	08-Jan-15	696	271	233	**-5%**
Period Average					**6%**