Pairs trading is a popular strategy among market neutral players. The paper reviewed today is “Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy” by Caldeira & Moura (2013).

I will give a short description of the strategy, paper and give my commentaries and backtest the strategy in python.

Pairs trading 101

The concept is simple: Two stocks are co-moving with a certain sensitivity. When the co-movement is broken, enter the position that benefits from stocks returning to the co-moving state. Let’s illustrate the case: I’ve hand - picked two obviously cointegrated assets: Visa and Mastercard.

[plot with visa+mastercard\*beta]

Look at the date X: if you short Visa and long Mastercard, you get a profit. There could be even reverse situations. Then we just have to enter a reverse position. Looks easy.

The strategy is meant to be market neutral. The trader only bets on the assumption that the cointegration remains after the divergence happens.

The paper:

Authors use a cointegration method, which was described by Vidyamurthy (2004). Be prepared: come math coming up. The method is based on guess what: Cointegration – a property of 2 time series that produce a stationary or normally distributed time series. It is mean reverting and thus fluctuates around 0 value.

First, we have to pick a right pair. To do that, the authors perform the following tests: Augmented Dickey Fuller, Engle Granger 2 step approach and Johansen test. When the pair is detected, spread between these stocks is extracted the following way: ,

The parameter is denoted as hedge ratio. Usually, it is calculated as a beta of regression of one asset on another.

After that, the authors normalize the time series, calculating it’s z-score.

The trading rule is to open a position when the z-score hits the 2 standard deviation thresholds from above or bellow. If the z-score hits the -2 standard deviation threshold, it means that the portfolio of pairs is bellow its long-run equilibrium value. In this case, one should buy the portfolio, which means buying stock l and selling stock s. If the z-score hits the 2 standard deviation threshold from above, the portfolio of pairs is overvalued and one should sell it short, which means selling stock l and buying stock s.

Buy to open if Sell to open if Close short position if Close long position if

zt < −2.00

zt > 2.00

zt < 0.75

zt > −0.50

If the ratio develops in an unfavorable way, we will use a stop-loss and close the position if we have a loss of 7%. Finally, we will never keep a position for more that 50 days

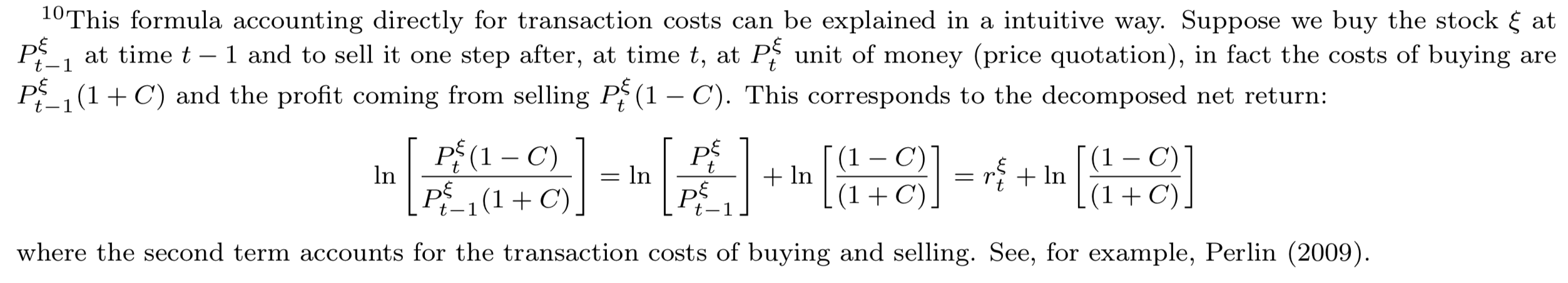
The returns on a pair is calculated with the following method:

After that they form a portfolio of 20 best performing pairs based on in-sample Sharpe ratios.

An important detail is that the training period constitutes 1 year and the testing period is 4 months. Once the testing period is finished, all positions are unwinded and the new iteration starts again.

Now let’s try to replicate the paper (at least to the extent that I understand the paper)

My notes on the paper and the method:

* Look at the beta calculation within this period and this: See the difference a couple of extra observartions make?
* Is training period defined for the tests to not to break the assumption of number>100? Ain’t it better to look at a more granular data
* 
* They claim that they enter and unwind the position each time! Investing or trading is usually compared to sex, but not in that fashion! They pull out too early!