Rise and Fall:
An Autoregressive Approach to Pairs Trading

Vasco Portilheiro (vascop@stanford.edu) and Bora Uyumazturk (yuyumaz@stanford.edu)

CS229

Goal

In finance, pairs trading is a strategy which bets that the spread between two similar stocks should be stable. The main idea is to identify pairs of stocks which are historically correlated, and then while trading, if they diverge from each other, bet that they will revert by shorting the higher one and buying the lower one. The simplicity of the idea has drawn many of the quantitative minds on Wall Street to tackle it. For our project, we tried to outperform a standard, non-learning strategy, from the finance literature by modeling the spread between stocks as an autoregressive process.

Data

For evaluation, we restricted our scope to only stocks listed on the S&P 500. We got our data from Yahoo Finance, which provided us with open and close prices at each day, as well as stock meta-information such as company name, symbol, and sector.

In our models, we looked at the normalized time series rather than the raw data. This simply means dividing the entire time series by the initial value, so that every stock begins at value 1.

Baseline Strategy

In the baseline model, taken from Gatev, et. al. (2006), we use a simple square distance metric to identify pairs. Given two time series, we denote the distance between them by:

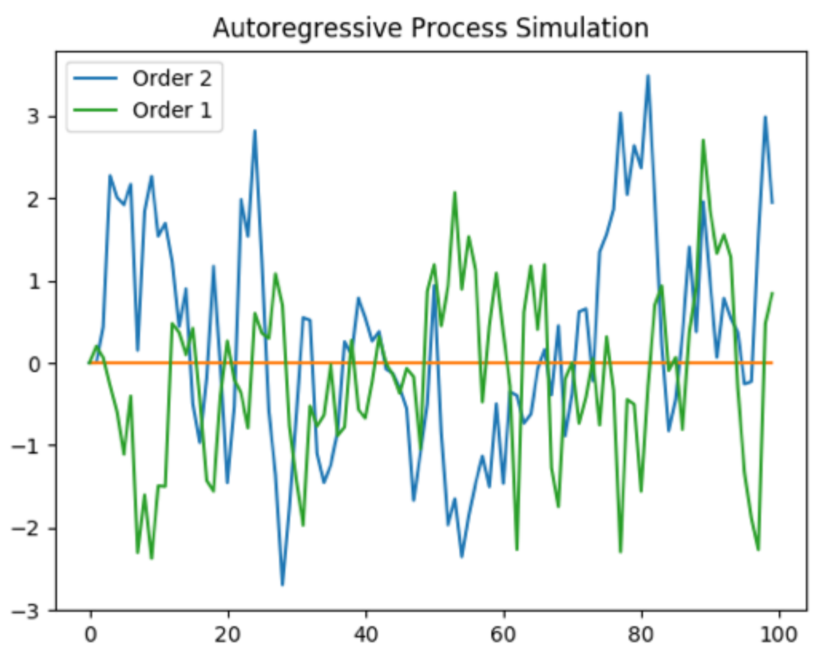
$$dist(X,Y) = \sum_t (X_t - Y_t)^2$$

During the trading period, we wait until the difference in spread is greater than two times the standard deviation over the course of the history before we enter a long-short position. Once they revert back to 0 spread, we exit the position.

Autoregressive Process

An autoregressive process is a random walk defined as follows:

$$X_n = \phi_0 + \sum_{i=1}^{p} \phi_i X_{n-i} + \varepsilon_n$$



Parameter Estimation

For each pair, we estimated the parameters of the p-th order autoregressive process governing their spread using maximum likelihood estimation. The log-likelihood function was as follows:

$$l(\phi,\sigma^2) = (n-1-p)\log(\frac{1}{\sqrt{2\pi\sigma^2}})$$
$$\frac{1}{2\sigma^2}\sum_{i=p+1}^{n}(X_i - \phi_0 - \sum_{j=1}\phi_j X_{i-j})$$

Defining our design matrix and labels as follows,

$$X = \begin{pmatrix} X_1 & X_2 & \dots & X_p \\ X_2 & X_3 & \dots & X_{p+1} \\ \vdots & \vdots & & \vdots \\ X_{n-p} & X_{n-p+1} & \dots & X_{n-1} \end{pmatrix} \qquad Y = \begin{pmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_n \end{pmatrix}$$

we found MLE estimate for the coefficients to be give by the normal equations:

$$\hat{\phi} = (X^T X)^{-1} X^T Y$$

and the variance of the noise to be:

$$\hat{\sigma^2} = \frac{1}{n-1-p}\sum_{i=p+1}^{n}(X_i - \phi_0 - \sum_{j=1}^{p}\phi_j X_{n-j})^2$$

Citations:

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.

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

Joshi, M.S. "The Concepts and Practice of Mathematical Finance", Cambridge University Press, (2008)

https://en.wikipedia.org/wiki/Autoregressive_model

Autoregressive Strategy

During the trading period, we only opened a trade at time t if the following two conditions were met:

$P(\text{convergence}|X_t) \geq 0.5$ and $P(X_{t+1} > X_t) < \tau$

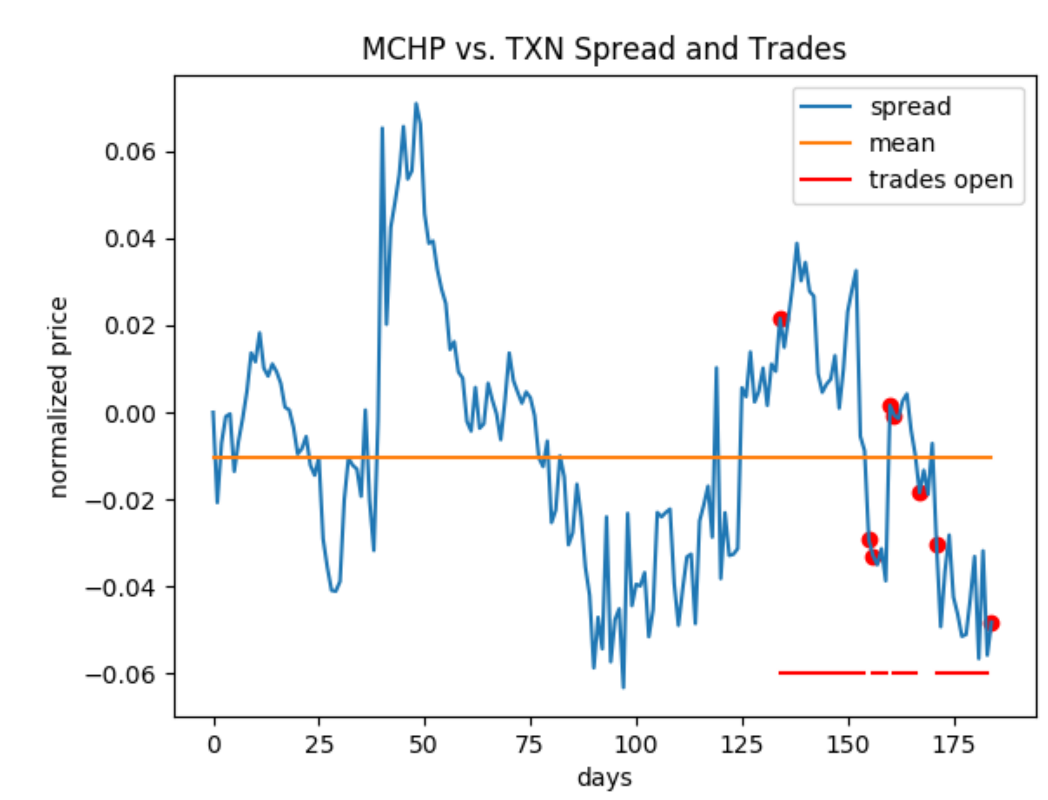
We call τ the 'growth threshold'. (We calculated the first quantity using Monte-Carlo simulation, and the second analytically)

Performance

Our test set came from 11/25/2013-11/25/2015, while our test set came from 11/25/2015 - 11/25/2017 (ROI here means return on investment). We averaged our returns over disjoint periods of 185 trading days, where we learn parameters for the first 125 days and trade for final 60.

Model	*Mean Pair ROI (Test)*	*Mean Pair ROI (Train)*
Base	*-0.0063*	*-0.0094*
AR(1), no growth threshold	*-0.00810*	*-0.0079*
AR(1), 0.2 growth threshold	*-0.0076*	*-0.0076*
AR(2), no growth threshold	*-0.00801*	*-0.0106*
AR(2), 0.2 growth threshold	*-0.0067*	*-0.0072*

Below we plot one pair traded with the AR(2) model with 0.2 growth threshold.



Next Steps

We have implemented a model which tracks the moving average so as to better handles shocks to the mean, and adapt to them. Preliminary testing has shown returns of over 1%.