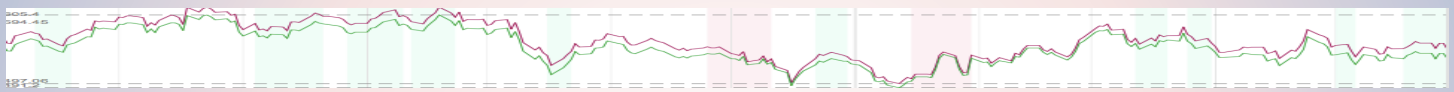
A PAIRS TRADING STRATEGY FOR GOOG/GOOGL USING MACHINE LEARNING

Jiayu Wu, jiayuwu@stanford.edu

Introduction

Pairs trading is a popular trading strategy in the last three decades after it was first used by Morgan Stanley in 1980s. Pairs trading means to utilize a pair or a bag of related financial instruments to make profits by exploiting their relations. One important feature of pairs trading is that it is market-neutral, which is particularly appealing in the current volatile and unpredictable macro-economic environments. In this project, we will use the spread model, the O-U mean-reverting model, and SVM to build a trading strategy and apply the strategy to GOOG/GOOGL. We will first illustrate the spread model and the O-U mean-reverting in detail. Unlike most previous work that only takes price spread into consideration, we will also use the spread model and the O-U mean-reverting model to model the two securities' technical indicators. In other words, we extend the concept of ``spread" by also investigating technical indicators' spread. We will construct trading signals by processing different kinds of ``spreads" and then use these trading signals as input features for SVM classification. Instead of using the traditional back-testing method to test our trading strategy, we will use SVM binary classification to measure our trading strategy. To achieve that, we will reconstruct the original pricing feeds to labeled examples, and there are two methods we use to reconstruct the labeled examples, one for measuring the strategy's ability to seize profit opportunities, and the other for measuring the strategy's ability to make directional predictions. One important thing for a pairs trading strategy is to select a proper pair of financial instrucments. For example, if the price of security A always rises when the price of security B rises, it seems that A and B may be used for pairs trading. However, the explicit relation between prices may not be good enough for a good pair. The good pairs should share as many the same intrinsic characteristics as possible. GOOG/GOOGL are both shares of Google Inc. (nos Alphabet Inc.) but with different vote rights. GOOGL represents Class A shares while GOOG represents Class C shares. Only Class A shares have voting rights. Therefore, generally, the price of GOOGL is slightly higher than that of GOOG. Other than voting rights, they are essentially the same since their prices are based upon the same fundamentals.



Models

We use the canonical spread model:

$$\frac{dA_t}{A_t} = \alpha dt + \beta \frac{dB_t}{B_t} + dX_t$$

Where A is the price of first security and B is the price of second security.

We are interested the mean-reverting residual term, which is dXt. Therefore, we use O-U process to model Xt:

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t$$

Trading Signal

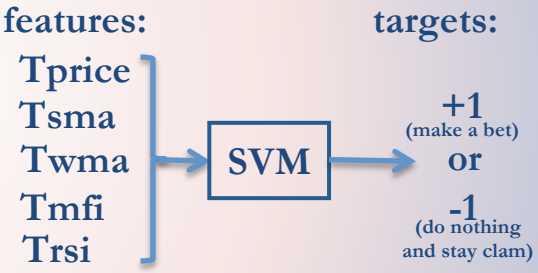
Here we define T-score for each feature:

$$T_{price} = \left|\frac{X_t^{price} - \mu_{price}}{\sigma_{eq}^{price}}\right|$$

By the definition, we can see that T-score is standardized version of X_t. We use the absolute value here because in this project we focus on the absolute value of the spread not its sign.

Likewise, we define the following T-scores for other technical indicators: $T_{sma} = \left|\frac{X_t^{sma} - \mu_{sma}}{\sigma_{eq}^{sma}}\right|, T_{wma} = \left|\frac{X_t^{wma} - \mu_{wma}}{\sigma_{eq}^{wma}}\right|, T_{mfi} = \left|\frac{X_t^{mfi} - \mu_{mfi}}{\sigma_{eq}^{mfi}}\right|, T_{rsi} = \left|\frac{X_t^{rsi} - \mu_{rsi}}{\sigma_{eq}^{rsi}}\right|.$

SVM



Data

We obtained data from Quantquote.com, 1-min feeds ranging from 10/01/2014 – 10/30/2015.
Data example: {'date': 20151030, 'time': 1559, 'open': 711.98, 'high': 712.58, 'low': 710.72, 'close': 710.78, 'volume': 41773}. Since we want to transform the return/profit prediction problem to a directional prediction/binary classification problem, we need to reconstruct the data set. Therefore, after reconstruction, a data example is: {Tprice: 0.0745, Tsma: 0.3250, Twma: 0.6684, Trsi: 0.3421, Tmfi: 1.837}

Results

Metrics 1:

	Positive	Negative	Total
Positive	$TP = 6745$	$FN = 3878$	10623
Negative	$FP = 2291$	$TN = 8222$	10513
Total	9036	12100	N

accuracy	precision	recall	F-measure	AUC
0.7081	0.7465	0.6349	1.0670	0.7085

Metrics 2:

	Positive	Negative	Total
Positive	$TP = 6993$	$FN = 3552$	10545
Negative	$FP = 2238$	$TN = 8353$	10591
Total	9231	11905	N

accuracy	precision	recall	F-measure	AUC
0.7261	0.7576	0.6632	1.0574	0.7259

Trading Strategy Framework

- **Step 1** •run linear regression to get spread model residuals
- **Step 2** •construct Xts using residuals from the previous step based on $X_t = \sum_{i=t_0}^{t} dX_t$
- **Step 3** •run lag 1 auto-regression on Xts to get parameters according to O-U model
- **Step 4** •use parameters obtained from the previous steps to compute T-scores for price and selected indicators
- **Step 5** •train SVM with the modified data set
- **Step 6** •apply the trained model to test data set, namely to make trading decisions, which is either "make a bet" or "stay calm and do nothing"