

MSBD 5013 Strategy Report

XIONG Di (20447802), TAN Rui (20460012), CHAI Di (20458239), WANG Guizheng (20459441)

Introduction

Quantitative Trading

Employing computer science and financial engineering to build investment models, quantitative trading assists investors in decision making and performs automated trading under the preset rules. Markowitz proposed mean-variance model in 1952, which introduced mathematic tools into the area of finance and made automated trading possible. Therefore, many asset management agency started to develop the quantitative trading strategies. For example, Morgan Stanley invented pair trading in 1985 and successfully made excess return within a short time. At the same time, stock and commodity exchanges gradually supported quantitative trading at the beginning of the millennium. From then on, quantitative trading strategies became popular among investors, because investors are desired to earn return as most as possible.

Our Project

A futures market is an auction market in which participants buy and sell commodity and futures contracts for delivery on a specified future date. In this project, we are provided with minute-level future data from China futures market, including 10 commodity futures and 3 stock index futures. Our strategy needs to trade every minute according to historical data and trend we predict.

In terms of future trading strategies, the most common are known as going long, going short and spreads. Going long means you expect the price of the commodity is going to increase. What we need to do is to predict the price of which commodity will increase in next days. Another way is going short, which is opposite to buying potentially increasing commodities. We can sell the contract at high price if we suppose this commodity will decrease. Spreads is a method of making profits of the price difference between two different contracts of the same commodity. We should buy a contract and sell it at higher price to get the net profit.

We developed two strategies in our project. The first one is a simple version of pair trading, the other is a strategy based on price prediction.

Data Preprocessing

There are 5 records for each future every minute in the given dataset, namely, open, high, low, close and volume. What we have used in data preprocessing is price related records. In order to choose the profitable futures to trade, we calculated the volatility of each future separately, and calculated the correlation coefficient between every two futures.

Volatility

$$Volatility_{i,j} = \frac{Price(High)_{i,j} - Price(Low)_{i,j}}{P(Avg)_{i,j}}$$

where i denotes future type, and j denotes trading days.

We measure the volatility of each future in a trading day. In other words, the highest price, lowest price and average price are derived from every 240 data records. Finally we calculated the average volatility for each future in all trading days.

$$\overline{Volatility_i} = \sum_j^n Volatility_{i,j}$$

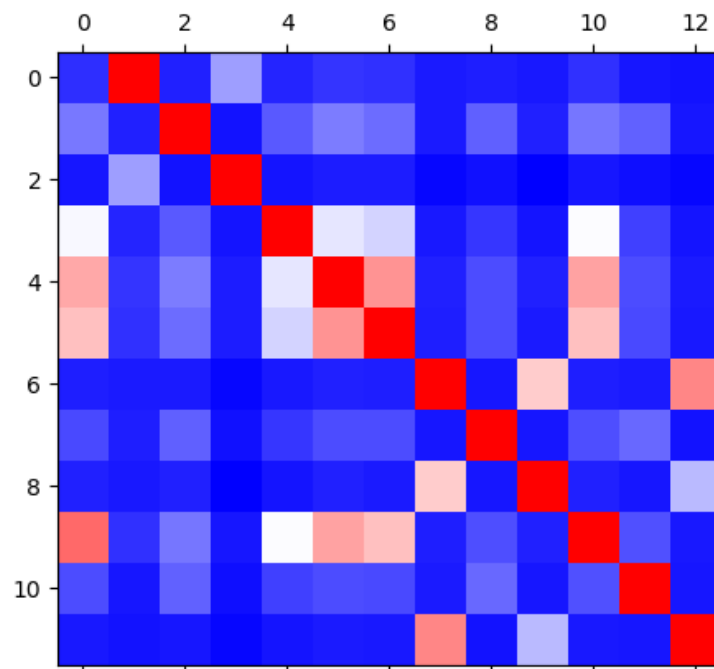
The futures with highest volatility are JM.DCE (0.018), I.DCE(0.017) and JM.DCE (0.015).

Correlation Coefficient

We calculated the difference of close and open as price changing for each future every minute. Then every future has a price changing vector. We used pearson correlation as the measure to determine similarity between two futures. Say one future's price changing vector is \mathbf{X} , while the other is \mathbf{Y} , then their correlation is calculated by

$$r(\mathbf{X}, \mathbf{Y}) = \frac{Cov(\mathbf{X}, \mathbf{Y})}{\sqrt{Var(\mathbf{X})Var(\mathbf{Y})}}$$

The figure below is a heatmap of correlation. The blocks in red color show that the corresponding two futures are highly correlated. One future itself, of course, is completely correlated, and we can see that there are still pairs in high correlation.



Pair Trading

Pair trading is a classic strategy, and it is the first strategy that we came up with. It assumes that the market is neutral and matches a long position with a short position in a pair of highly correlated items. The strategy waits for deviation of the correlation, then long on the under-performing item while short on the over-performing item.

Find Trading Pair

It turned out by our data analysis that coke and cole are the most correlated item pair among all combination of 13 futures. So we chose coke and cole as the pair that we trade with.

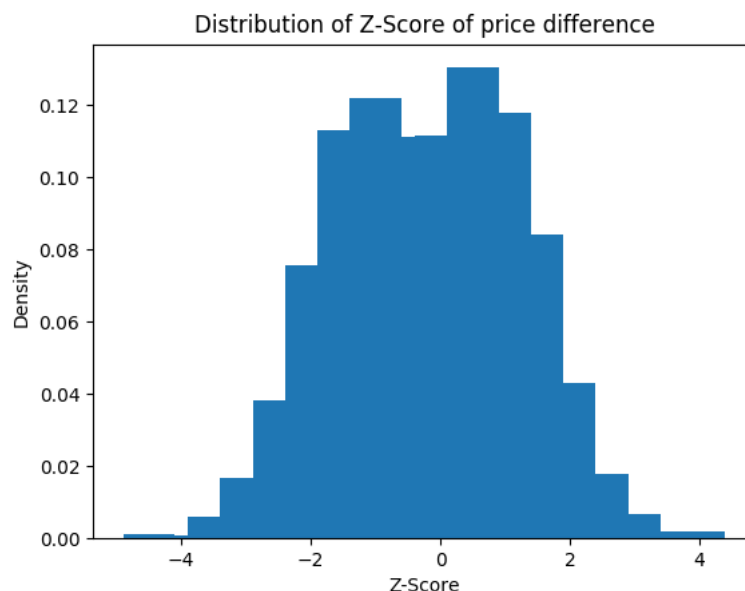
(complete the analysis)

Setting Parameters

In pair trading, if the two items are highly correlated, then their price difference will remain in a rather stable state. If the price difference is 'abnormal', then the strategy is triggered and long the under-performing item and short the over-performing item. Now, the problem is, how to define 'abnormal'? We built a sequence of price difference in the prior 30 minutes, compute the mean and variance of the sequence, then derive the current zscore.

$$zscore = \frac{current - mean}{variance}$$

After observing at the distribution of zscore, we found out that the probability of $zscore > 3$ is under 5%. So we set 3 as the trigger , and 0.5 as the close out threshold.



Moreover, we set a stop loss parameter of this strategy. It records the highest total balance and compare current balance with it. If current balance is lower than 90% of highest total balance, the strategy is forced to close out. In other words, we used max draw down as the stop loss parameter, and the max draw down is 10%.

Performance

We used all data in the past four months to test the strategy, since it doesn't need any training. It has a cumulative return of 18.5% (annual return 54%), and its sharpe ratio is 1.71. The max draw down is -9%, which approximates our preset rate.

The Price Predictor Method

Feature engineering is an important part in machine learning. The intuitive way to forecast the futures' price is simply using the historical price and volume as input. We assumed that there are some latent patterns in futures' price behavior. Maybe one of the patterns is that, if the futures price goes down at first and increases in the pass few minutes, then the futures price will

go up in the few minutes. This is a simple fluctuation process, but with periodic pattern. But we don't know whether it is true or not. And practically, we can't identify all the patterns. In this situation, machine learning can help a lot.

To solve this problem by machine learning, we can first generate features which represents the futures' price behavior. Then set the label to positive if the price increases and negative if the price decreases.

After these two steps, we transferred the price prediction problem into a binary classification problem.

Feature Engineering

Firstly, we want to predict the futures price in the next 10 minutes. So the label is +1 if the close price after 10 minutes is higher than current, otherwise the label is 0.

And we chose the data in the pass 10 minutes to generate the feature. The formula is like:

$$Feature_1 = (Close - Open)_{past-10mins}$$

$$Feature_2 = (High - Close)_{past-10mins}$$

$$Volume = (volume)_{past-10mins}$$

We combined these three features above and get a thirty-dimensional feature.

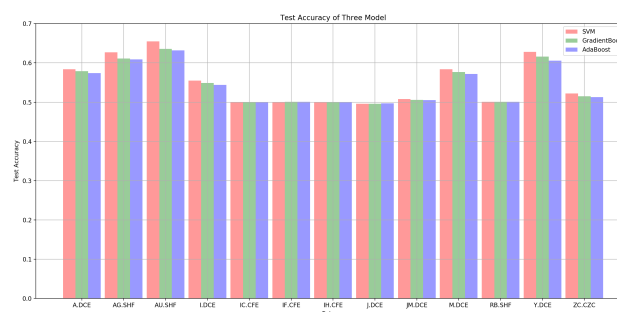
We also used the ATR(Average True Range) and MACD(Moving Average Convergence/Divergence) to analysis the futures price and get a four-dimensional feature (the output of MACD is macd, macdsignal, macdhist).

Model

We used SVM, AdaBoost, GradientBoost methods to do the prediction.

The kernel function of SVM is RBF. AdaBoost and GradientBoost are two ensemble methods, and we used 50 week classifiers in practice.

The result of these three methods on each futures is like:



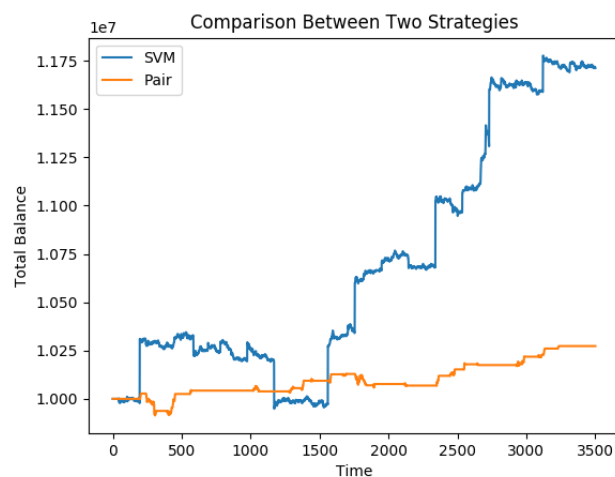
The figure shows that the prediction of 6 futures are much more higher than 50%, which means our feature is working and the models generated some latent patterns. We can also see that SVM is better than the other two models.

Performance

From the figure above, we chose AU.SHF as the item we trade with in this strategy since it has the highest prediction accuracy. Data from 2017-07-17 to 2017-11-17 are used for training, and we ran backtest on data from 2017-11-20 to 2017-12-15. The cumulative returns is 15.6%, sharpe ratio is 3.78 and max draw down is -7.6%.

Conclusion

After running backtests on the dataset, we found that our methods have excess return with a rather low max draw down. The pair trading method is simple and effective, but it's performance is not distinguishing. It's likely that the market is now in a much higher trading frequency now, and the price difference fluctuation are discovered quickly. So we need to make more rules to improve the adaptation of the simple strategy.



For the machine learning method, we should know that the dataset given is quite small. And robust quantitative trading strategies are selected after running backtests on years of data. A small dataset would easily lead to overfitting and downgrade the performance. And for improvement, we can engineer more features to identify special patterns and make the strategy more robust.

We have posted our code on github. Please click the link to see more details.

<https://github.com/lovingzw/helicopter>