Project Proposal

Title: Deep Reinforcement Learning for Multi-Dimensional Pairs Trading in the Brazilian stock market

Student: Tiago Monteiro Cardoso Rio de Janeiro, October 1st, 2018

1. Historical Background

Pairs Trading is an old statistical arbitrage trading strategy (1) that uses asset pairs whose historical prices are correlated. Since in the short term the correlation is usually weak, whenever the spread between the assets differs from its equilibrium level, an arbitrage opportunity appears: the most expensive asset is sold, the cheapest asset is bought and the position is undone once the spread value returns to a predetermined distance. Due to the proprietary nature of the strategy, academic interest for pairs trading has been scant (2). As a matter of fact, pioneering works which were able to establish the strategy's profitability have only recently been published (3, 4). In Brazil, particularly, a fundamental condition for the practice of financial arbitrage - securities lending - has only been regulated in 1996 (5) and the main academic studies (6, 2) have been published in the last decade (2009 and 2013).

In addition, an expansion of the strategy - "Multi-dimensional pairs trading" or MDPT - has been recently proposed (7). In MDPT the strategy includes more than two stocks. Although the multi-dimensional "pairs" promise better accuracy in predicting arbitrage opportunities, the strategy's promise of higher profits (8) is still in need of further experimental evidence. As far as the employed techniques are concerned, there has been much documented use of reinforcement learning in the algorithmic trading domain, but to date the sole proposed model-free reinforcement learning agent for pairs trading so far published is the one described in (9).

To sum up, the relevance of the present project lies in the fact that for the first time a deep reinforcement learning agent will be implemented for conducting a multi-dimensional pairs automated trading strategy. As a bonus, I hope to be able to demonstrate the profitability of the approach for the Brazilian market.

2. Problem's description

In this project I tackle the problem of devising a reinforcement learning agent to execute an automated multi-dimensional pairs trading strategy. Particularly, the agent will be trained on historical records of quotations of three highly correlated stocks from the Brazilian stock market. After the training period, the agent will start to issue coupled orders: to buy the undervalued stock and sell the overvalued ones or, vice-versa, to sell the overvalued one and buy the other two (another possible action will be to maintain the current position). In the simulation we will begin with a portfolio consisting of zero shares and a given amount of money. At the end of each trading day the portfolio value will be recalculated. At the end of the whole testing period, total returns will be computed and compared to two benchmarks.

3. The data

Input data consists of daily quotations (from 01/Jan/2000 to 01/Oct/2018) of three stocks listed in the Brazilian stock exchange - ITSA4 (Itaúsa - Investimentos Itaú S.A. preference shares), ITUB3 (Itaú Unibanco Holding S.A.) and ITUB4 (Itaú Unibanco Holding S.A. preference shares), together with the values of the Ibovespa index. The agent will be trained and tested in this dataset in order to learn to make buying/selling decisions involving the assets. As for the Ibovespa index, it will be used to construct one of the benchmarks against which the agent's performance will be compared. The information in the dataset is public and was retrieved from <finance.yahoo.com> in csv format. Four files were downloaded, three corresponding to the stocks and one to the Ibovespa index. In the files each row of data corresponds to one trading day and there are seven columns: date, the share's open value, the highest value, the lowest value, the close value adjusted for splits, the close value adjusted for both dividends and splits and the volume of shares negotiated. The files were retrieved on 02/Oct/2018 and are sent together with the present proposal for revision.

4. The solution

The goal of this project is to build a reinforcement learning agent to optimize decisions regarding the opening and closing of transactions following a recently proposed multi-dimensional pairs trading scheme. The agent will control a portfolio comprising three stocks and will follow the strategy outlined in (7). According to that strategy the agent will buy/sell the stock which is under/overvalued relative to the other two and will simultaneously sell/buy the other two stocks. Consequently there are in all 6 different possible operations and, in order that the strategy should remain market risk-free, the value allocated to the short position (sell) in each operation must equal the value allocated to the long position (buy). Since there is one "operation" more (i.e. neither buying nor selling), there are a total of 7 actions for our agent. The reward will be the daily return. The agent will learn through the maximization of the long-term discounted accumulated returns.

5. Benchmark

At the end of the testing period, total returns will be computed and we will check whether the agent was able to outperform one or both the chosen benchmarks - the returns calculated with the Ibovespa index and the returns of a random agent. The Ibovespa is the financial index of the Brazilian stock exchange - Bovespa - and is the standard benchmark against which Brazilian investments are evaluated. For the second benchmark, an agent will be implemented to choose randomly and with equal probability from the available actions.

6. Evaluation metrics

In order to evaluate our agent, we will compute, for the period under study, the portfolio's return rate obtained by the agent and compare it to the return rates obtained by the benchmarks. The return rate is given by the formula:

$RR = \frac{P_f}{P_i} - 1$, where RR is the return rate, P_i is the initial portfolio value and P_f is the portfolio value at the end of the testing period.

7. Project's design

GENERAL FRAMEWORK AND STRATEGIES

This project concerns a RL agent with actor and critic networks and a replay memory. At each step (trading day) the simulator will render available a new state corresponding to the stocks' quotations (both the current ones and those from a number of previous days). There will be 7 possible actions and the reward will be the daily return. The actor network will learn a policy function $\pi(a|s, \theta_P)$, where π is stochastic, that is, its outputs are probabilities assigned to the possible actions, and θ_P, the network parameters, are learned through gradient ascent with the goal of maximizing the long-term accumulated returns (discounted by a factor gamma). As for the critic network, it will estimate a value function $V^{\pi}(s_t)$ whose objective is the minimization of the mean-squared-error between the state's estimated value and its target value under policy π.

ALGORITHM

The algorithm to be implemented is the one suggested in (10).

ASSUMPTIONS AND SIMPLIFICATIONS

We plan to use a policy gradient framework, which is especially appropriate for the continuous action space of the general case (trading whichever amount yields a higher return). However, for the moment being I decided to solve the simpler problem corresponding to a discrete action space. Thus, in each simulation step, the agent is only allowed to trade the same value (100 units of Brazilian currency for the asset being bought (sold) and 50 units for each of the other two assets, which are being sold (bought)), and consequently, there are only 7 actions available (as described in 4). We will assume that transactions are free of costs and that the leverage ratio is not restricted, that is, there is no limit for loans. We also assume that a real market agent can make transactions using the adjusted close price.

8. References

(1) Göncü A, Akyildirim E, *A stochastic model for commodity pairs trading*. *Quant. Finance*, 2016, **16**(12), 1843-1857.

(2) Caldeira J, Moura, G V, *Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy*. Available at SSRN: https://ssrn.com/abstract=2196391 or http://dx.doi.org/10.2139/ssrn.2196391.

(3) Gatev E, Goetzmann W N, Rouwenhorst K G, *Pairs trading: Performance of a relative-value arbitrage rule*. Review of Financial Studies 19(3):797-827 (2006).

(4) Do B, Faff R (2010). *Does simple pairs trading still work?*, Financial Analysts Journal 66(4): 83-95.

(5) Takimoto E, *A estratégia pairs trading no mercado de ações brasileiro*. Dissertação de Mestrado, 2007, Faculdade IBMEC, São Paulo.

(6) Perlin M, *Evaluation of pairs-trading strategy at the Brazilian financial market*. *J Deriv Hedge Funds*, 2009, **15**(2), 122-136.

(7) Lau C A, Xie W, Wu Y. *Multi-Dimensional Pairs Trading Using Copulas*. Retrieved from < http://www.efmaefm.org/0EFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2016-Switzerland/papers/EFMA2016_0390_fullpaper.pdf> in 02/10/2018.

(8) Xie W, Liew R Q, Wu Y, Zou Xi. *Pairs Trading with Copulas*. The Journal of Trading 11:3, 41-52. 2016.

(9) Hwang T, Norris S, Su H, Wu Z, Zhao Y. *Deep Reinforcement Learning for Pairs Trading*. Retrieved from < http://manul.io/img/gekkos/report.pdf> in 02/10/2018.

(10) Shen Y, Zhao Y. *Deep Reinforcement Learning for Pairs Trading Using Actor-critic*. Retrieved from < > in 04/10/2018.