1
Introduction

Stock market prediction is an issue of increasing interest for both investors and researchers. In recent years, it has been widely studied using different approaches, including Artificial Intelligence (AI) techniques [1]. Despite these efforts, accurately predicting future stock trends and developing business strategies capable of turning good predictions into profits in the real world are still big challenges [2].

The main difficulty in predicting stock market prices is due to the nonlinearity and the large amount of uncertainty and noise shown by stock market data sources [3, 4, 5].

These characteristics imply that classical statistical predicting methods do not accurately predict stock market time series. These time series require more powerful methods [6], such as, Partial Least Squares Regression [7], Support Vector Machines [8, 9, 10] and Artificial Neural Networks [2, 5, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21].

This work focuses on stock market time series forecasting. We present a model for predicting stock market behavior that aims to guide the investor both into Pairs Trading and buy and sell operations.

Initially proposed by Gerald Bamberger in the late 80's, Pairs Trading refers to a neutral market operation [22, 23]. It is basically an arbitrage transaction between two assets. This famous operation uses the relationship between two assets, that shows a correlated price variation pattern. This usually happens when both assets represent the same company or reflect the same economical fundaments. The spread is defined as the ratio between the prices of two assets. When two related assets vary similarly in percentage, their spread does not change. Hence, any pronounced change in the spread suggests a good opportunity to simultaneously buy one of the assets and sell the other.

We explore two different forecasting schemes. In the interday forecast, only interday data is used. In the intraday forecast, in addition to input features used in our first forecasting scheme, we also include the intraday data already known at prediction time, aiming at reducing the daily forecasting error [7].

In both forecasting schemes, we use three regression tools as predictor algorithms, which are: Partial Least Squares Regression (PLSR), Support Vector Regression (SVR) and Artificial Neural Networks (ANN).

The choice for ANN and SVR has been mainly motivated by their capacity of finding patterns in nonlinear data [4, 14, 15, 24] and its ability to easily deal with irregularities [13, 25], and uncertain, incomplete or insufficient data [19].

In an attempt to also achieve good predictions with a linear model, we try the PLSR method. As shown by the results in the experiments, our attempts are successful with some data preprocessing.

Recent publications using similar models show good results for stock market predictions [1]. Nevertheless, their use and evaluation in real world scenarios by taking advantage of predictions to get investment profits, is still understudied. The main and most frequent reason for this is that the proposed systems are evaluated by using classical prediction error metrics, such as Mean Percentage Error (MPE) or Mean Absolute Percentage Error (MAPE). These metrics do not capture the system impact in market operations results, but do give us an idea of the model generalization power [2].

The best way to deal with this deficiency is to use a trading system to assess the prediction quality. This is our goal. Hence, to evaluate the predictions, we harness the trading system capacity to simulate an investor's decision making process in the stock market. In this context, a trading system is one that operates according to a fixed set of rules and contains no discretionary components [5].

It is important to note that the system performs two autonomous tasks, thus the predictor and the trading system work independently. Once the predictor has completed the forecasting, the trading system is able to assess the prediction quality. This evaluation is performed by simulating the trader actions from predictions to good profits. The simulation follows the predefined rules and takes into account the real stock market constraints, such as brokerage commission rates, slippage, stock exchange costs and income tax. It also considers possibilities such as short selling and stock rent.

Therefore, more than using three different predictors for stock market forecasting, we propose a trading system capable of using the predictions to find recommended times of the day when either to buy or sell a particular asset or to perform Pairs Trading. The major advantages of using a trading system are the removal of psychological and emotional characters of the decision making process, the risk reduction through the definition of a risk control mechanism aiming to avoid unwaited losses, and the obtainment of operational metrics closer to the ultimate investor's goal: to use forecasts to build profitable trade strategies.

The proposed trading system carries out only day trading. An operation is called day trading when all the practice of buying and selling stocks are performed in pairs and within the same trading day. In this case, all positions are closed until the moment when the market closes for trading and the investor has no stocks in his portfolio after the market closing. Day trading is especially interesting in crisis periods, when the high daily volatility gives investors a great opportunity to raise profits, even though this scenario usually exposes them to a greater risk [2].

In this work, we build predictors to foresee the current day minimum and maximum values. If we consider buy and sell operations, these values correspond to the price of the asset to be traded. If the focus is on Pairs Trading, these are spread values between the considered asset pair. Next, we use their predictions to guide the investor's decision making. Hence, the trading system can assess the forecasting quality. To the best of our knowledge, the combination of predictors with a Pairs Trading negotiation model is novel. Most of the previously proposed systems have been only applied to the usual buy and sell operations.

Our system is tested by simulating both types of trades in the BM&FBOVESPA Stock Exchange, the official Brazilian stock exchange. Due to its market value, it is the world's third largest stock exchange [26]. It is also a very appealing financial market, since it presents a significant increase both at the number of individual investors and at its daily average trading volume, as shown in figures 1.1 and 1.2, respectively. However, there is a lack of studies examining its behaviour.

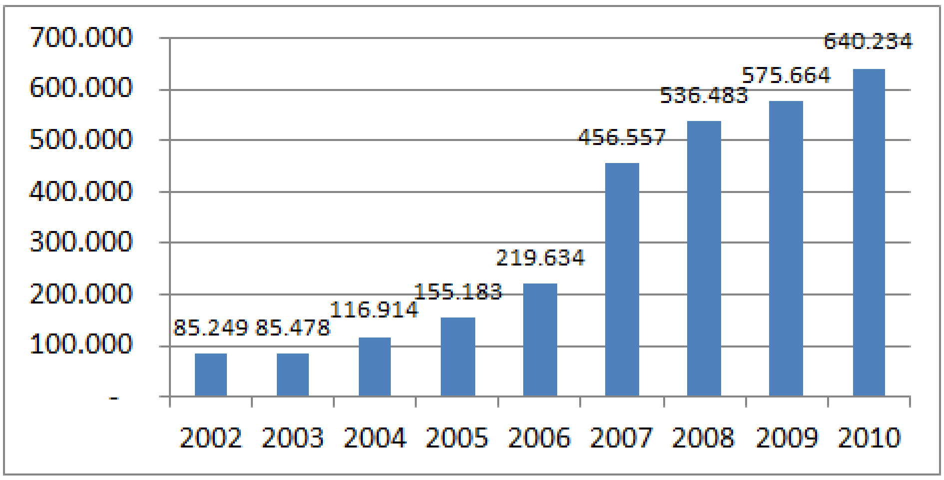

Figure 1.1: Number of Individual Investors in BM&FBOVESPA

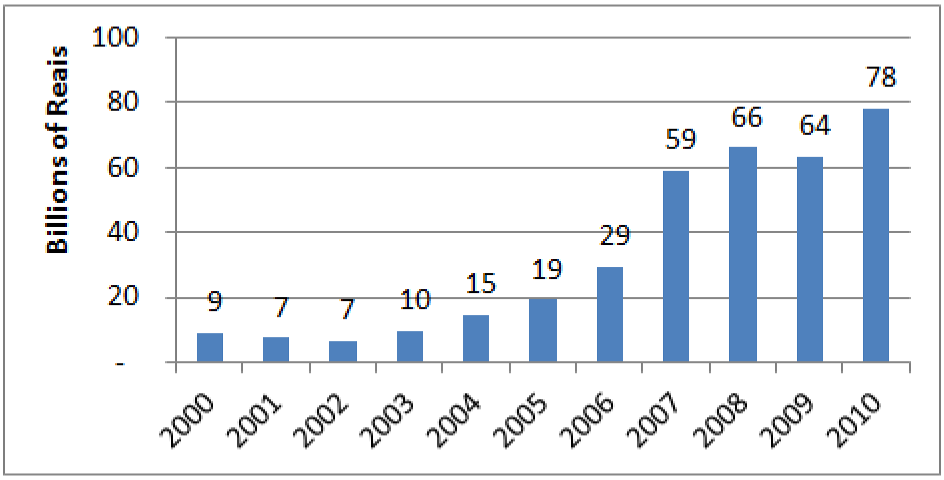

Figure 1.2: Daily Average Trading Volume of BM&FBOVESPA

In our experiments considering buy and sell operations, we examine assets of the most heavily traded Brazilian companies. And to simulate Pairs Trading, we use six pairs where some kind of relationship can be easily observed.

For benchmarking, we compare our predictors to four other predictors: one naive predictor and three more elaborate predictors proposed by O'Connor and Madden [27]. Furthermore, in order to examine how close our predictors are to the optimal solution, we present the predictions of an Oracle which acts by simulating an optimal predictor that performs the forecasting task accurately.

The dataset period falls between December 18^th, 2008 and May 14^th, 2010, which corresponds to 343 trading days. It is worth mentioning that this period includes a serious international financial crisis. Even so, the results obtained by the proposed trading system with using the predictions performed by our predictors, show a very high profitability. The yield in some cases amounts to an annual return on investment of more than 300%.

The programming language chosen for the development of the whole system was C++. And experiments were run on a Pentium Core 2-Duo 3.2 ghz.

This dissertation is structured as follows. In Chapter 2, we briefly discuss related findings that compose the state-of-the-art. In Chapter 3, we further describe the forecasting task, the proposed predictors for both interday and intraday approaches and the dataset used, as well as how it is organized according to the type of trade considered. In Chapter 4, we explain the simulation task and how our trading system works in the performed tests. Both buy and sell, and Pairs Trading experiments are presented in Chapter 5, where we depict how the feature engineering is conducted and report the quality assessment for both forecasting and simulation tasks. Finally, we present in Chapter 6 some conclusions and directions for future work.