Optimizing Stock Market Price Prediction using a Hybrid Approach Based on HP Filter and Support Vector Regression

Meryem Ouahilal[1], Mohammed El Mohajir[2], Mohamed chahhou[2], Badr Eddine El Mohajir[1]
[1]*Faculty of Science, Abdelmalek Essaadi University*
Tetuan, Morocco
m.ouahilal@gmail.com
b.elmohajir@ieee.ma
[2]*LIMS, Faculty of Science Dhar El Mehraz, Sidi Mohamed Ben Abdallah University*
Fez, Morocco
m.elmohajir@ieee.ma
mchahhou@hotmail.com

***Abstract*—Predicting stock prices is an important task of financial time series forecasting, which is of great interest to stock investors, stock traders and applied researchers. Many machine learning techniques have been used in recent times to predict the stock price, including regression algorithms which can be useful tools to provide good accuracy of financial time series forecasting. In this paper, we propose a novel hybrid approach which combines Support Vector Regression and Hodrick-Prescott filter in order to optimize the prediction of stock price. To assess the performance of this proposed approach, we have conducted several experiments using Maroc Telecom (IAM) financial time series. It is daily data collected during the period between 2004 and 2016. The experimental results confirm that the proposed model is more powerful in term of predicting stock prices.**

Keywords—Stock price prediction, Time series forecasting, Support vector regression, Hodrick-Prescott filter, Decision support.

I. Introduction

In the stock market, the closing price is the final price at which a security is traded on a given trading day. The closing price represents the most up-to-date valuation of a security until trading commences again on the next trading day. The closing prices provide a useful marker for investors to assess changes in stock prices over time - the closing price of one day can be compared to the previous closing price in order to measure market sentiment for a given security over a trading day.

An economic time series consists of multiple components corresponding to short-term irregular and seasonal variations, a medium-term business cycle, and a long-term trend movement. Most macroeconomic analysis is concerned with a medium-term business cycle and a long-term trend movement. However these fundamental movements are hidden in the original economic data because various irregular and seasonal variations are dominant in the data [1] [2]. Therefore it is often difficult to read directly from the original data the fundamental movement of an economic variable under study.

The financial time series includes some noise that may influence the information of the dataset. For better understanding and analysis of the trend, and optimize the accuracy of stock price prediction, noise filtering is necessary before using the prediction model.

There have been many studies using machine learning techniques to predict the stock price. A large number of successful applications have shown that regression algorithms can be very useful tools for financial time-series modelling and forecasting.

Our objective in this research work is to propose a hybrid approach that combines support vector regression (SVR) and Hodrick-Prescott filter (HP), for enhancing the prediction of stock price by learning the historical data of IAM using our proposed model.

The rest of this paper is organized as follow. In the section II we present the Hodrick-Prescott filter and how to use it for noise filtering. Section III provides a brief theoretical overview of the SVR model. Section IV presents our methodology. Section V focuses on prediction of stock market prices using our hybrid model, and we then present and discuss the experimental results of our case study. Finally, section VI concludes this work.

II. Hodrick-Prescott for Noise Filtering

It is inevitable for a time series to include some noise that may influence the whole information of the dataset. In the stock market, the volume of stocks fluctuate every day and do not show any signs for forecasting in the stock market, thus resulting in difficulty to understand the trend of the change in it.

However, for a macro perspective of the stock market, the long-term trend should be predicted and discovered. Although this long-term trend cannot give an obvious indication which specific stock will rise tomorrow, it reveals nonetheless the performance of the whole investment environment and to a certain extent gives important hints helping make decisions on the stock market [3] [4] [5].

To better understand and analyze the trend, noise filtering is essential. In this research, we will use Hodrick-Prescott filter to analyze the data and realize noise filtering with it.

The Hodrick–Prescott filter is a mathematical tool used in macroeconomics, especially in real business cycle theory, to remove the cyclical component of a time series from raw data.

It is used to obtain a smoothed-curve representation of a time series, one that is more sensitive to long-term than to short-term fluctuations. The adjustment of the sensitivity of the trend to short-term fluctuations is achieved by modifying a multiplier λ.

The goal of HP filter is to decompose the time series into several series with common frequencies. Let y_t be our data at time-period t. we want to decompose the data into growth component, τ_t, and the cyclical component, c_t

$$y_t = \tau_t + c_t \qquad \text{for } t=1,\ldots,T \tag{1}$$

One attraction of the HP filter is that it may be applied to nonstationary time series (series containing one or more unit roots in their autoregressive representation), a relevant concern for many macroeconomic and financial time series.

The HP filter removes a smooth trend τ_t from a time series x_t by solving the minimization problem

$$min \sum_{t=1}^{T}[(x_t - \tau_t)^2 + \lambda((\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1}))^2] \tag{2}$$

With respect to τ_t . The residual, or deviation from trend $z_t=x_t-\tau_t$ is commonly referred to as the business cycle component, and is the object of economic interest. In this sense the HP filter is a highpass filter, removing the trend and returning high-frequency components in z_t.

The parameter λ penalizes fluctuations in the second differences of x_t, and must be specified by the user of the HP filter.

III. SUPPORT VECTOR REGRESSION

A. Support Vector Machines

In machine learning, support vector machines (SVM) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.

More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin).

When support vector machines were used to solve the regression problem they were usually called support vector regression (SVR) [6] [7].

B. Support Vector Regression

SVR performs linear regression in the high-dimension space using ε –insensitive loss and, at the same time, tries to reduce model complexity by minimizing$\|\omega\|^2$.

This can be described by introducing (non-negative) slack variables ξ_i, ξ_i^* $i = 1, \ldots n$, to measure the deviation of training samples outside ε –insensitive zone.

Thus SVR is formulated as minimization of the following functional:

$$\frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{3}$$

$$min \begin{cases} y_i - f(x_i, \omega) \le \varepsilon + \xi_i^* \\ f(x_i, \omega) - y_i \le \varepsilon + \xi_i \\ \xi_i, \xi_i^* \ge 0, i = 1, \ldots n \end{cases} \tag{4}$$

This optimization problem can be transformed into the dual problem and its solution is given by

$$f(x) = \sum_{i=1}^{n_{sv}}(\alpha_i - \alpha_i^*)K(x_i, x) \quad 0 \le \alpha_i^* \le C, 0 \le \alpha_i \le C \tag{5}$$

Where n_{sv} is the number of support vectors (SVs).

C. Kernel Function

Kernel function should satisfy the Mercer's Theorem. Four common kernel functions include:

- Linear function: $K(X_k, X_l) = \langle X_k | X_l \rangle$;
- Polynomial function with parameter d: $K(X_k, X_l) = (\langle X_k | X_l + 1 \rangle)^d$;
- Radial Basis Function (RBF) with parameters β: $K(X_k, X_l) = \exp(-\beta\|X_k - X_l\|)^2$;
- Hyperbolic tangent: $(X_k, X_l) = \tanh(2\langle X_k | X_l \rangle + 1)$.

IV. METHODOLOGY

In this paper, we propose a novel hybrid approach based on the combination of the Hodrick-Prescott filter (HP) and the Support Vector Regression algorithm (SVR). So we propose a new model of financial time series prediction based on our hybrid approach.

The objective of this approach is to improve and optimize SVR model predictions with the help of HP filter that will parse and normalize our data by filter and remove all existing noise in our financial time series.

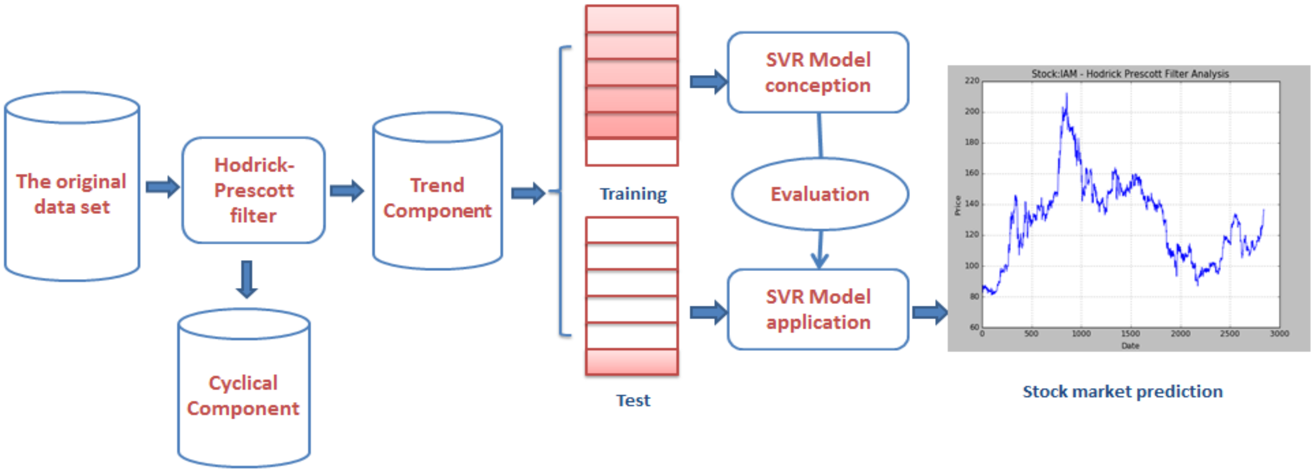

Figure 1: Model of stock price prediction using SVR model and HP filter

V. STOCK MARKET PREDICTION

A. DataSet description

In the stock market, the closing price is the final price at which a security is traded on a given trading day. The closing price represents the most up-to-date valuation of a security until trading commences again on the next trading day.

The closing prices provide a useful marker for investors to assess changes in stock prices over time - the closing price of one day can be compared to the previous closing price in order to measure market sentiment for a given security over a trading day [8].

Thus, the closing price is selected as our prediction target of the original data set. It is daily data collected by IAM during the period between 2004 and 2016.

Our data set has 6 attributes and 2840 samples. They are Date, Open price, Close price, High price, Low price, and Volume. The goal is to predict Close price for different amount of time in the future.

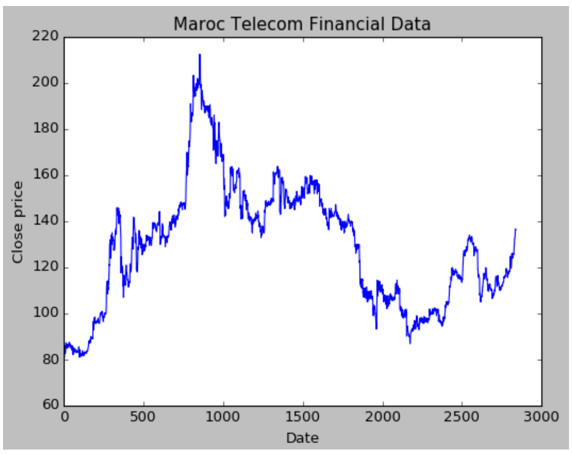

Figure 2: Maroc Telecom (IAM) financial time series

B. Our hybrid approach

The regression analysis focuses on the Close price on the *(t+1)-th* day changes when the Open price, Close price, High price, Low price and Volume on the *i-th* day vary.
Our goal is to fit the following relationship by regression analysis:

$$Close_{t+1} = f\,(Open_t\,,\; Close_t\,,\; High_t\,,\; Low_t\,,\; Volume_t)$$

Before performing the regression, we need to use Hodrick-Prescott filter to filter noise and normalize the data value on each attribute separately.

The goal of HP filter is to decompose the time series into several series with common frequencies. We want to decompose the data into the trend and the cyclical components.

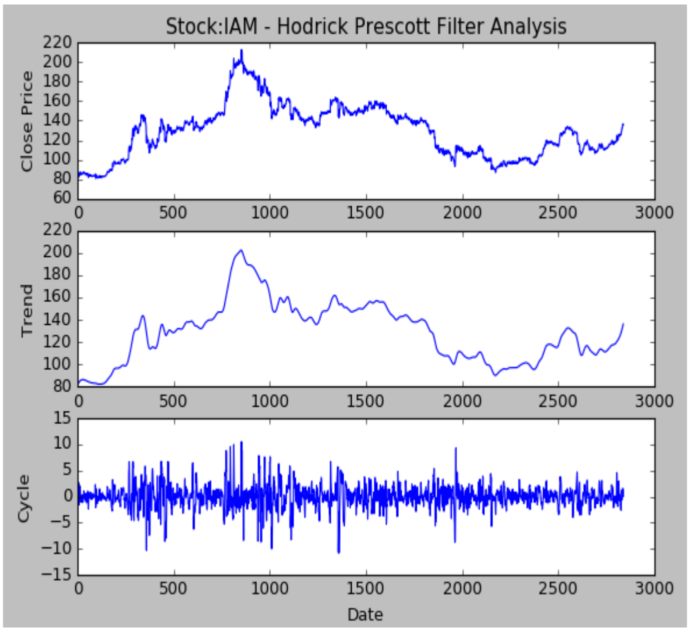

Figure 3: Hodrick-Prescott filter analysis

Above chart shows the Stock IAM price and HP Filter components trend and cycle component. It's clearly visible that trend component is ultra-smooth and very good in predicting the future of IAM price direction. And the Cycle Component extreme values suggest a possible trend reversal.

C. Experimental results

The regression performances vary with different selection of four important parameters: (1) the kernel function (2) penalty parameter *c* (3) kernel parameter *g* and (4) degree of the kernel function *d*. Four-stage grid search is used to find the best combination of parameters.

In prediction experiments, the data are divided into two subsets. The data from December 2004 to December 2012 was employed as training set used for training the models of the algorithms. We have selected four folds of testing set decomposed as follow:

- The data from January 2013 to December 2013 are employed as first fold of testing set.
- The data from January 2013 to December 2014 are employed as second fold of testing set.
- The data from January 2013 to December 2015 are employed as third fold of testing set.
- The data from January 2013 to July 2016 are employed as fourth fold of testing set.

The final parameters selected for this case are kernel='poly', c= 275 and *d*=3.

Figure 4, 5, and 6 show the results of our regression by plotting the original data and regressive data together for different amount of time in the future.

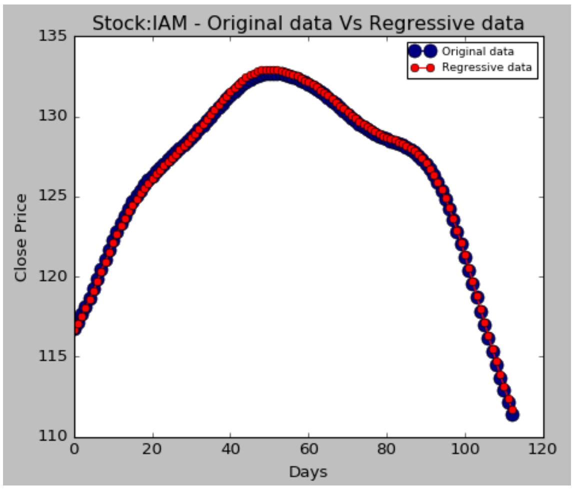

Figure 4: The prediction of stock price for one year

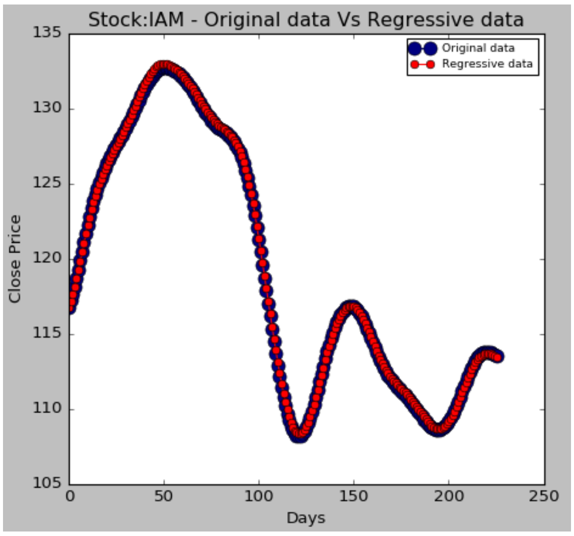

Figure 5: The prediction of stock price for two years

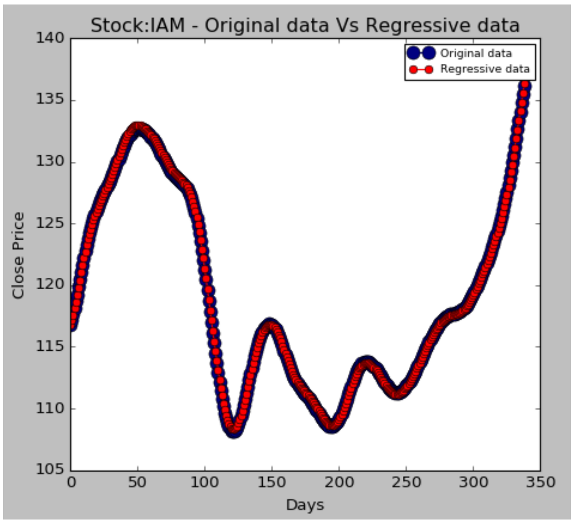

Figure 6: The prediction of stock price for three years

The error rate is computed between the actual and predicted stock prices come from the experiments. To calculate the error rate, Mean average percentage error (MAPE) is used in this study.

It's defined in the following:

$$MAPE(y, y') = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - y'_i|}{|y_i|} \times 100\% \qquad (6)$$

Where *y'* and *y* represent the predicted result and observed value respectively and *N* is the sum of training samples [9].

Figure 7 and Table 1 show the error (average MAPE) committed in different SVR models using different kinds of filters. The calculation of MAPE was done for the testing dataset and the training dataset.

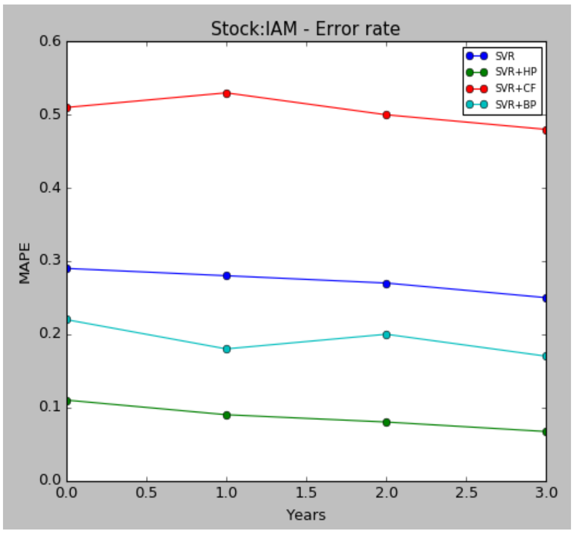

Figure 7: MAPE (Error rate) of the predictive models

Table 1: MAPE (Error rate) of the predictive models

Years	SVR	SVR+HP	SVR+CF	SVR+BP
2013	0.29	0.11	0.51	0.22
2013-2014	0.28	0.09	0.53	0.18
2013-2015	0.27	0.08	0.50	0.20
2013-2016	0.25	0.067	0.48	0.17

Where HP is the Hodrick-Prescott filter, CF is the Christiano Fitzgerald filter and BP is the Band-Pass filter using the Fourier transformation. These filters are the most known and used in financial time series analysis [10].

D. Discussion

To assess the performance of our proposed approach we have conducted several experiments for predicting stock price with different SVR models using different kinds of filters.

The objective was to verify that the combination of SVR model and HP filter provide the better results of stock price prediction compared to others filters.

Effectively, the combination of SVR model and HP filter provide the better results since the MAPE error given by this model is the lowest among all proposed error rate.

According to our experimental results, we can conclude that the proposed model using our hybrid approach with combine the SVR model and the HP filter is a powerful predictive tool for stock price and financial time series.

VI. Conclusion

Predicting stock price is a major factor in stock market prediction and has been paid much attention. Therefore, the applications of regression model in financial field are a meaningful attempt.

In this research work, we proposed a novel hybrid approach of predicting stock price which combines Support Vector Machine model and Hodrick-Prescott filter in order to optimize the stock price prediction.

To assess the performance of this proposed approach, several experiments have been conducted using Maroc Telecom (IAM) financial time series.

The experimental results have shown that the proposed model gives better results in term of stock price predictions.

REFERENCES

[1] B. Krollner, B. Vanstone and G. Finnie, Financial time series forecasting with machine learning techniques: A survey, Information technology papers, April 30, 2010

[2] M. Higo and S. Kuroda Nakada, How can we extract a fundamental trend from an economic time series? , Discussion Paper No. 98-E-5

[3] Robert M. de Jong and N. Sakarya, The Econometrics of the Hodrick-Prescott filter, September 22, 2013

[4] M. Deistler and C. Zinner, Forecasting Financial Time Series, Canberra, February, 2007

[5] Peter C. B. Phillips and S. Jin, Business Cycles, Trend Elimination, and the HP Filter, June 2015

[6] P. Meesad and R. Islam Rasel, Predicting Stock Market Price Using Support Vector Regression, 2013

[7] H. Huang, W. Zhang, G. Deng and J.Chen, Predicting Stock Trend Using Fourier Transform And Support Vector Regression, 2014 IEEE 17th International Conference on Computational Science and Engineering

[8] R.Varshney and A.Mojsilovic, Business Analytics based on financial time series, IEEE Signal Processing Magazine Volume 28, Issue 5, September 2011

[9] M. Tiwari, M. Bhai Jai and O. Yadav, Performance analysis of Data Mining algorithms in Weka, IOSR Journal of Computer Engineering Volume 6, Issue 3, PP 32-41, 2012

[10] Ruey S. Tsay, Analysis of Financial Time Series, Wiley Series In Probability And Statistics, 2005