Support Vector Machine and Least Square Support Vector Machine Stock Forecasting Models

Lucas Lai*, James Liu

1Computer Department, University of Polytechnic, Hong Kong, China
*Corresponding Author: lucaskclai@yahoo.com

Copyright © 2014 Horizon Research Publishing All rights reserved.

Abstract This paper explores the Support Vector Machine and Least Square Support Vector Machine models in stock forecasting. Three prevailing forecasting techniques - General Autoregressive Conditional Heteroskedasticity (GARCH), Support Vector Regression (SVR) and Least Square Support Vector Machine (LSSVM) are combined with the wavelet kernel to form three novel algorithms Wavelet-based GARCH (WL_GARCH), Wavelet-based SVR (WL_SVR) and Wavelet-based Least Square Support Vector Machine (WL_LSSVM) to solve the non-linear and non-parametric financial time series problem. This paper presents a platform for comparison of the wavelet-based algorithm using Hang Sang Index, Dow Jones and Shanghai Composite Index which has significant influence to each other. It has been discovered that wavelet-based model is not as good as the LS-SVM model. The best result is from LS-SVM without wavelet-based kernel.

Keywords Autoregressive Conditional Heteroskedasticity; Support Vector Regression; Least Square Support Vector Machine; Wavelet Transform; Daubechies waveletes; Symlet Wavelets

1. Introduction

The argument over the practical use of Artificial Intelligence to forecast financial market is a very sensitive and controversial issue. In the book of [4], it says that the prices of securities fully reflect available information in the Efficient Market Hypothesis (EMH). Investors buying securities in an efficient market should be expected to obtain an equilibrium rate of return. Weak-form EMH asserts that stock prices already reflect all information contained in the history of past prices. The semistrong-form hypothesis asserts that stock prices already reflect all publicly available information. The strong-form hypothesis asserts that stock prices reflect all relevant information. Under EMH, it is possible to extract information from the historical prices of the stock, as an input to the forecasting tools to project the future value. The argument here is almost everyone particularly the securities player will access to different forecasting news. Once the news is available, the market will digest the impact of the news and there is no advantages in using it hence it will become useless. However, the issue here is the accuracy of the forecasting news and which form of EMH is the target market. USA stock market is a typical strong-form EMH while Chinese stock market is a weak-form EMH. Moody, Standard and Poor ratings and their forecast news are very popular but how often we rely on their forecast news are very popular but how often we rely on their forecast news to trade!

Reference [3] explained that US equity returns have been predictable for many years especially in the long run. Earnings yieldshas had clear empirical advantages over dividend yields. Earnings yields is the benchmark on how well the company performs while dividend yields is the ability of the company to distribute its profit. It is not always a good indictor as banking and utilities sectors have steady dividend yield while new Initial Public Offering (IPO) will not be so generous. The use of dividend yield as a predictive variable leads to a basis in forecasting regression. [31] proved that random walk is not a sufficient and necessary condition for EMH. [30] found out that Chinese stock market cannot be classified as weak-form EMH. [9] proved that the β parameter of a company (which is a ratio between stock returns and market moves did not show significant relationship. Capital Asset Pricing Model - CAPM is based on market portfolio but in reality it is difficult to find. [15] stated that CAPM is not applicable to recent Chinese stock market. He also mentioned CAPM is robust but Arbitrage Pricing Theory (APT) easily analyses all factors affecting the stock price. The proof of CAPM is rigid but not APT. In 1992, using NYSE, AMEX and NASDAQ he found out β has nothing to do with the company size. All these findings using modern investment theories could be confusing as it is difficult to draw conclusion on how to use it. This is probably because the market is not easy to be defined and there is no single market that would not be affected by others. Today’s economic model is quite different from that 10 or may be 20 years ago and it would make the financial forecast even more challenging. It is necessary to develop new tools and methodologies in financial forecast as the markets are becoming more robust and complicated.
The objective of this paper is to review the wavelet-based forecasting models through which we would like to test the predictability of the models and compare those without the wavelet-based models. The models are based on GARCH, SVR and LSSVM. They are set to forecast the actual daily close value of Hong Kong Hang Sang Index (HSI) given the past 5-year records. HSI has been selected because it reflects the semi-strong-form EMH [2]. Hong Kong being the third largest financial trading centre cannot be compared with the US market which has a very long history, enormous trading volume, pioneer of financial reform and impeccable securities law. Before Hong Kong was a follower of the US market until recently that Chinese market has significant impact on it. Hong Kong investment advisor [27] has pointed out that the Hong Kong Stock market is not efficient and lack of volume like the US stock market to support the development of other approaches like artificial intelligence method. His theory will be challenged and this paper has shown that the proposed models can accurately predict Hong Kong Stock market using the latest forecasting techniques. [7] forecasted the volatility of stock index and [18] predicted the stock returns which are an indirect approach for the actual index value. The actual index value from these approaches may not be useful. It is well known throughout the literature that financial time series particularly stock index is non-linear. The three main factors of such time series are trend, seasonal and stochastic. These 3 factors affect the prediction result in stock index as it is impossible to develop a model to integrate all these factors. [13] used Chaotic Oscillatory-based Neural Networks and Lee Oscillator to successfully catch the variability period of HSI between 2007 and 2008. But it was a pattern prediction rather than actual value forecast. The application of the stochastic factor in stock forecast is limited, hence we focus on the trend and season and our challenge is to find out the best model for the prediction task. Despite the fact that stock index forecast has been conducted for many decades, the latest artificial intelligence techniques such as GARCH, SVR and LSSVM have improved the degree of prediction accuracy. Our objective is to seek for the best algorithm from the current techniques and apply it to recent financial time series.

This paper explores the prediction performances of wavelet-based models such as WL_GARCH, WL_SVRand WL_LSSVM in predicting exact stock prices on the Hang Sang Index (HSI) over a 4-day and 20-day forecasting horizons respectively. There are 5 trading days in a week but wavelet-based models can only deal with even number of days and hence a 4-day cycle is chosen to represent a week. In order to compare the 4-day short-term forecast, a 20-day long-term forecast is selected which is 4 weeks to represent a month. The model will give a 4-day and a 20-day ahead forecast respectively. In addition, the same datasets were employed in GARCH, SVR and LSSVM without the wavelet-based kernel as comparison. This paper is an extension of the work from [14] on using SVR in stock forecast but wavelet-based kernel is introduced. SVR was conducted with the software system from [16], LSSVM was conducted using the LS-SVMLAB toolbox which is provided by KatholiekeUniversiteit Leuven[26] while the experiment of GARCH was conducted with MATLAB GARCH toolbox. The three wavelet-based algorithms, WL_GARCH, WL_SVRand WL_LSSVM, are developed by the authors under MATLAB environment using GARCH, SVR and LSSVM as the basic kernel.

The rest of this paper is organized as follows. Section 2 describes the method, description on GARCH, SVR, LSSVM and Wavelet Transform function. Section 3 provides the empirical modeling of our models and the empirical result. Section 4 gives the conclusion and outlines our future work.

2. Methods

Three different markets DJ, HSI and SH historical data are input into the above 6 forecasting models. It is the objective of this paper to test the accuracy of the forecasting result using hybrid kernel-based function.

2.1. GARCH

GARCH (General Autoregressive Conditional Heteroskedasticity) by Bollerslev is a linear time series prediction method. It is a standard textbook material in econometrics and finance[6]. There are many families of GARCH as described in [11] and its application is throughout the financial institutes. GARCH models are designed to capture certain characteristics that are commonly associated with financial time series such as fat tails, volatility clustering leverage effects. One branch of GARCH called Ngarch as described in [22] is an alternative approach to the famous Black Scholes Model. ARFIMA-FIGARCH from[25]that can predict the Indian Stock Data during the period 3 July, 1990 to 18 September 2009 accurately. In [12]paper, GARCH prediction on NK225 has the RMSE value of 0.2013 while that of the pure SVM is 0.1820 and the best RMSE value from Wavelet-based RVM is 0.0202.

2.2. Support Vector Regression

The following is a brief description on SVR for nonlinear function estimation such as the financial times series. In the primal weight space the model takes the form

\[ f(x) = \omega^T \phi(x) + b, \quad \text{(1)} \]

with the given training data \( \{x_k, y_k\}_{k=1}^N \) and \( \phi(.) : R^n \rightarrow R^m \) a mapping to a high dimensional feature space which can be infinite dimensional and is only implicitly defined. Note that in this nonlinear case the vector \( \omega \) can also become infinite dimensional. The optimization problem in the primal weight space becomes

\[ \min_{\omega, \xi, \xi^*} J_\rho (\omega, \xi, \xi^*) = \frac{1}{2} \omega^T \omega + C \sum_{j=1}^N (\xi_j + \xi_j^*) \quad \text{(2)} \]
subject to:

\[ y_k - \omega^T \varphi(x_k) - b \leq \varepsilon + \xi_k, \quad k = 1, ..., N \]

& \hspace{1cm}
\[ \omega^T \varphi(x_k) + b - y_k \leq \varepsilon + \xi_k^*, \quad k = 1, ..., N \]

\[ \xi_k, \xi_k^* \geq 0, \quad k = 1, ..., N. \]

Applying the Lagrangian and conditions for optimality, the following is the dual problem

\[ \max_{\alpha, \alpha^*} J_D(\alpha, \alpha^*) = \frac{1}{2} \sum_{k=1}^{N} (\alpha_k - \alpha_k^*) \sum_{k=1}^{N} (\alpha_k - \alpha_k^*) - \sum_{k=1}^{N} \gamma_k (\alpha_k - \alpha_k^*) \]

subject to:

\[ \sum_{k=1}^{N} (\alpha_k - \alpha_k^*) = 0 \]

\[ \alpha_k, \alpha_k^* \in [0, c] \]

Here the kernel trick has been applied with

\[ K(x_k, x_l) = \varphi(x_k)^T \varphi(x_l) \] for \( k, l = 1, ..., N \). The dual representation of the model becomes

\[ f(x) = \sum_{k=1}^{N} (\alpha_k - \alpha_k^*) K(x_k, x) + b \]

Consider the following Vapnik’s \( \varepsilon \)-insensitive loss function

\[ L_\varepsilon(y - f(x)) = \begin{cases} 0, & \text{if } |y - f(x)| \leq \varepsilon \\ L(y - f(x)) - \varepsilon, & \text{otherwise} \end{cases} \]

Eq. 5 is a convex cost function where \( L(\cdot) \) is convex. Primal problem

\[ \min_{w, b, \alpha, \alpha^*} \frac{1}{2} w^T w + C \sum_{k=1}^{N} (L(\xi_k) + L(\xi_k^*)) \]

subject to

\[ y_k - w^T \varphi(x_k) - b \leq \varepsilon + \xi_k \]

\[ w^T \varphi(x_k) + b - y_k \leq \varepsilon + \xi_k^* \]

\[ \xi_k, \xi_k^* \geq 0 \]

where \( \xi_k, \xi_k^* \) are slack variables. Here, \( x_k \) is mapped to a higher dimensional space by the function \( \varphi \) and \( \xi_k \) is the upper training error (\( \xi_k^* \) is the lower) subject to the \( \varepsilon \)-insensitive tube \( |y_k - w^T \varphi(x_k) - b| \leq \varepsilon \). The parameters which control the regression quality are the cost of error \( C \), the width of the tube \( \varepsilon \), and the mapping function \( \varphi \).

The constraints imply that we should put most data \( x_k \) in the tube \( |y_k - w^T \varphi(x_k) - b| \leq \varepsilon \). If \( x_k \) is not in the tube, there is an error \( \xi_k \) or \( \xi_k^* \) which we must minimize the objective function \( \text{SVR} \) to avoid under-fitting or over-fitting the training data by minimizing the training error \( C \sum_{k=1}^{N} (L(\xi_k) + L(\xi_k^*)) \) as well as the regularization term \( \frac{1}{2} w^T w \).

The Lagrangian for this problem is

\[ L(\omega, b, \varepsilon, \alpha, \alpha^*, \eta, \eta^*) = \frac{1}{2} \omega^T \omega + \sum_{k=1}^{N} (L(\xi_k) + L(\xi_k^*)) \]

subject to:

\[ \sum_{k=1}^{N} \alpha_k (\varepsilon + \xi_k - y_k - \omega^T \varphi(x_k) + b) - \sum_{k=1}^{N} (\eta_k \varepsilon_k + \eta_k^* \varepsilon_k^*) \]

With Lagrange multipliers \( \alpha_k, \alpha_k^*, \eta_k, \eta_k^* \geq 0 \)

So far from (1) to (8), \( \text{SVR} \) estimation function combined with the loss function is the foundation of the \( \text{SVM} \).

Support Vector Machine (SVM) is used in many machine learning tasks such as pattern recognition, object classification, and with regression analysis in time series prediction in Support Vector Regression, or \( \text{SVM} \), a methodology in which a function is estimated using observed data which in turn is used to train the SVM. It differs from traditional time series prediction methodologies in that there is no model in the strict sense – the data drives the prediction. [19] used \( \text{SVM} \) to determine the minimum enclosing zone and [10] used \( \text{SVM} \) in predicting country investment risk.

\( \text{SVM} \) has been used in long term stock market forecasting. [21] used an accelerated Levenberg-Marquardt algorithm to predict the stock market series of the Jakarta Stock Indices over 10 months, achieving an RMSE of 1.96%. [2] applied \( \text{SVM} \) to forecast the price trend for a single Chinese stock. [20] used \( \text{SVM} \) to predict the first day returns of US stock market IPOs, but found to be accurate in only 18% of cases. [29] claimed a profit over two months using a methodology that combined news and technical indicators. [12] used \( \text{SVM} \) to forecast the direction of stock movements which was correct 73% of the time. [24] reported the use of \( \text{SVM} \) in financial time series prediction over a 5-day forecasting horizon.

2.3. Least Square Support Vector Machine

LSSVM regression is closely related to regularization networks, Gaussian processes to reproduce kernel Hilbert spaces but emphasizes primal-dual interpretations in the context of constrained optimization problems. It is relatively a new tool, there is very little research in financial forecasting using LSSVM such as [28].
The following is a brief description of LSSVM mechanism on regression problems. Given a training data set \( \{x_k, y_k\}_{k=1}^{N} \), we can formulate the following optimization problem in the primal weight space:

\[
\min_{w,b,e} J_p(w,e) = \frac{1}{2} w^T \omega + C \sum_{k=1}^{N} e_k^2
\]

such that \( y_k = w^T \phi(x_k) + b + e_k \), \( k = 1, ..., N \) is modified here at two points comparing with (1) from the SVR section above. First, instead of inequality constraints one takes equality constraints where the value \( y \) at the left hand side is rather considered as a target value than a threshold value. Upon this target value an error variable \( e_k \) is allowed such that misclassifications can be tolerated in the case of overlapping distributions. These error variables play a similar role as the slack variables in SVR. Secondly, a squarer loss function \( e_k^2 \) is taken for this error variable. These modifications will greatly simplify the problem.

### 2.4. Wavelet Transform

The wavelet transform (WT) has been found to be particularly useful for analyzing signals which can best be described as aperiodic, noisy, intermittent and transient [1]. It really began in the mid-1980s where they were developed to interrogate seismic signals. The application of wavelet transform analysis in science and engineering really began to take off at the beginning of the 1990s. WT and Fourier transform (FT) are very similar in nature especially FT has been around since the 1800s. FT is built from sines and cosines functions which are periodic waves that continue forever. This approach is only good for signals that have time-independent wave-like features, signals which have more localized features for which sines and cosines do not model very well. WT is a different set of building blocks to model these types of signals [2]. In this paper, WT is tested if it can improve the forecasting accuracy of financial time series. It was invented by the Hungarian mathematician Alfred Haar. The most commonly used set of DWT was formulated by the Belgian mathematician Ingrid Daubechies in 1988 which is one of the methods considered in this paper. This formulation is based on the use of recurrence relations to generate progressively finer discrete samplings of an implicit mother wavelet function; each resolution is twice that of the previous scale. There are a number of families in Daubechies and Haar is the first one. Daubechies wavelets are quite asymmetric, in order to improve symmetry while retaining simplicity, Daubechies proposed Symmlets as a modification to her original wavelets (also symlets). The Daubechies and Symmlets wavelets are employed here in this paper.

Reference [23] described the conventional factor model, the data-generating process of each variable is the sum of two components: a component associated with factors common to all series and an idiosyncratic term. The underlying idea is that one can summarize the large information set into a small number of variables, the common factors, which retain the main features. Wavelet multi-resolution analysis allows one to decompose a time series into a low-frequency base scale and higher-frequency scale. Those frequency components can be analyzed individually or compared across variables. A). Times series are decomposed to orthogonal components of different frequencies. B). Each time scale uses a model to fit in. C). Overall forecast is obtained by recombining the components. [23] only used Symlet wavelet at level 4. Here, we used Symlet wavelet functions with coefficients from 2 to 8 and Daubechies wavelet function coefficients from 1 to 20 for comparison. The selection of such coefficients are based on the work [14].

The discrete wavelet transform (DWT) can be written as:

\[
T_{m,n} = \int_{-\infty}^{\infty} x(t) \psi_{m,n}(t) dt
\]

where the integers \( m \) and \( n \) control the wavelet dilation and translation respectively. By choosing an orthonormal wavelet basis, \( \psi_{m,n}(t) \), we can reconstruct the original signal in terms of the wavelet coefficients, \( T_{m,n} \), using the inverse discrete wavelet transform as follows:

\[
x(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} T_{m,n} \psi_{m,n}(t)
\]

The orthonormal discrete wavelets are associated with scaling functions and their dilation equations as follows:

\[
\phi_{m,n} = 2^{-m/2} \phi(2^{-m} t - n)
\]

They have the property

\[
\int_{-\infty}^{\infty} \phi_{0,0}(t) dt = 1
\]
The scaling function can be convolved with the signal to produce approximation coefficients as follows:

\[ S_{m,n} = \int_{-\infty}^{\infty} x(t) \phi_{m,n}(t) dt \]  

(14)

We can represent a signal \( x(t) \) with a combined series expansion using both the approximation coefficients and the wavelet coefficients as follows:

\[ x(t) = \sum_{n=\infty}^{\infty} S_{m_0,n} \phi_{m_0,n}(t) + \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} T_{m,n} \psi_{m,n}(t) \]  

(15)

3. Empirical Modeling

3.1. Data

The objective of this paper is to predict the 4-day and 20-day horizons of HSI closing value given the historical data of HSI. Based on the winner of ENUNITE [17], the benchmark to measure the forecasting accuracy is the mean absolute percentage error and our aim is set that to below 2. The historical data of HSI during August 2003 till June 2009 is downloaded from Yahoo financial website and it is separated into two datasets. The first set during 5 July 2007 till 30 June 2009 with 488 records is used to predict the 4-day with a sliding window of 248 days which is roughly a one-year dataset. The first shift-window during 5 July 2007 till 8 July 2008 is used to predict the next 4-day from 9 July 2008 onward. The next shift-window during 11 July 2007 till 14 July 2008 is used to predict the next 4-day from 15 July 2008 onward. Totally, there are 60 results. Another set, during 18 August 2003 till 30 June 2009 with 1448 records is used to predict the 20-day with a sliding window of 248 days. The first shift-window during 18 August 2003 till 16 August 2004 is used to predict the next 20-day from 17 August 2004 onward. The next shift-window during 16 September 2003 till 13 September 2004 is used to predict the next 20-day from 14 September 2004 onward. Totally, there are 60 results. The above data range is a test on the model robustness to highly volatile market as it ended near the financial tsunami. As a summary, a one year sliding window of 248 days is applied to the 488 records (5.7.2007-30.6.2009) to predict the stock price in the next 4 days, and to the 1448 records (18.8.2003-30.6.2009) in order to predict the stock price in the next 20 days. The purpose is to test the general forecasting ability of each model.

Using the same methodologies, two sets of index values of Shanghai composite Index and Dow Jones Index with the same record length and roughly the same period (Shanghai composite index 17.7.2003-30.6.2009 & 3.7.2007-30.6.2009 and Dow Jones 30.9.2003-30.6.2009 & 25.7.2007-30.6.2009) were analyzed by these models. As mentioned in the introduction, Shanghai composite index – China stock market is a weak-form EMH, HSI – Hong Kong stock market is semi-strong-form EMH and Dow Jones Index – US stock market is a strong-form EMH. Our purpose is to put these 3 markets to test under the above models and hypothesis that the strong form EMH should perform better than weak form of EMH. It also provides a foundation that our models can handle all kinds of market and its robustness in handling extreme data values during financial tsunami. The unprecedented financial tsunami is once in a life time experience for all financial institutions to handle. Compared with the last financial crisis in 1997 due to the collapse of Long Term Capital Management, the magnitude is far greater. The following figures are the characteristics of these data range.
nonlinear regressor. The last elements of the resulting matrix will contain the future values of the time series, the others will contain the past inputs. The following is a simple example.

\[
A = \begin{bmatrix}
    a_1 & a_2 & a_3 \\
    b_1 & b_2 & b_3 \\
    c_1 & c_2 & c_3 \\
    d_1 & d_2 & d_3 \\
    e_1 & e_2 & e_3 \\
    f_1 & f_2 & f_3 \\
    g_1 & g_2 & g_3
\end{bmatrix}
\]

\[W = \text{windowize}(A, [1 2 3])\]

Windowize is the relative index of data points in matrix \(A\), that are selected to make a window. Each window is put in a row of matrix \(W\). The matrix \(W\) contains as many rows as there are different windows selected in \(A\). It has been discovered this method outperforms the RDP as it is easier to apply. \([14]\) employed RDP5, RDP10, RDP15 and RDP20 to perform the same function as the windowize.

### 3.2. Forecasting Models and Parameters

Six algorithms have been developed in this paper. There are parameters in each model that require the algorithm to search in order to get the best result. \(C\) parameters are set to 500, 1,000, 5,000, 10,000, 20,000, 40,000 and \(g\) set to 1, 2 for the SVR and WL_SVR model based on the work of \([14]\). \(C\) is the value in (1) and \(g\) is the parameter of the mapping function \(\phi\). For the wavelet-based kernel, discrete wavelet transform is used and two types of methods are employed. The first is Daubechies with coefficients from 1 to 20 and the other is Symlet with coefficients from 2 to 8.

### 3.3. Empirical Results

\[
\text{MAPE} = 100 \sum_{i=1}^{n} \left| \frac{A - P}{A} \right| \frac{n}{m}
\]

MAPE stands for Mean Absolute Percentage Error which is the measure of accuracy in a fitted time series value in statistics, specifically trending. \(A\) and \(P\) are the real and the predicted values of the close value of the HSI respectively and \(n\) is the time frame or number of days.

| Table 1. Empirical Result in forecasting Hang Sang Index expressed in MAPE |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Forecast Horizon | 4 days          | 20 days         | 4 days          | 20 days         | 4 days          | 20 days         | 4 days          | 20 days         |
| SVR             | 0.4937          | 1.8217          | 0.4037          | 1.9353          | 0.6787          | 1.2924          | 0.4037          | 1.8291          |
| WL_db_svm       | 3.2709          | 9.2519          | 2.1003          | 4.4247          | 1.2170          | 2.1459          | 1.3914          | 2.1887          |
| WL_sym_svm      | 1.4682          | 10.1778         | 2.1571          | 4.4247          | 0.5890          | 2.9880          | 2.1503          | 2.7954          |
| LSSVM           | 0.8372          | 1.1407          | 0.8397          | 6.1239          | 1.7730          | 2.7961          | 0.8101          | 100             |
| WL_db_lssvm     | 1.4045          | 3.7927          | 1.7167          | 1.7666          | 1.4045          | 3.7927          | 1.7167          | 1.7666          |
| WL_sym_lssvm    | 2.1936          | 2.6297          | 1.9534          | 1.7368          | 2.1936          | 2.6297          | 1.9534          | 1.7368          |
| WL_db_garch     | 3.3246          | 6.6324          | 2.1701          | 1.5146          | 2.2104          | 1.5966          | 0.5192          | 1.8205          |
| WL_sym_garch    | 2.1939          | 5.237           | 2.1542          | 1.7941          | 2.3885          | 1.6668          | 0.5192          | 1.8267          |
From Table 1, SVR has 4 best MAPE, wavelet transform models has 3 while LSSVM has only 1. It seems the winner is SVR model. In general wavelet transform has improved the accuracy in GARCH models except the data range 2006-2011. However, the application of wavelet transform to SVR and LSSVM do not produce the same result. Most likely, it is because SVR and LSSVM use windowize method to pre-process the data and then map data into higher dimension. But in WL_db_svm, WL_sym_svm, WL_db_lssvm and WL_sym_lssvm, windowize method cannot be applied to the transformed data from the wavelet functions but only use normalization. For GARCH, log return method is used while its wavelet models use normalization. It seems SVR and LSSVM are more robust because of the windowize technique. In each of the basic model SVR, LSSVM and GARCH, wavelet based transform models have improved the accuracy. This confirms that the application of wavelet based models in previous work has significant improvement on financial time series forecasting. The best MAPE result is from the short-term 4-day forecast. In fact, it all comes from SVR model. For the long-term 20-day forecast, each model has its merit. The best result is 0.4037 4-day forecast horizon of 2011 from the above table. In Lin (2001) ENUNITE competition, he won the best forecasting result with MAPE 1.9. It is our target to keep the MAPE accuracy in the use of wavelet functions only happen in GARCH model. The sum of the best result of the 4-day and 20-day MAPE for Shanghai Composite Index is 5.3249, HSI is 5.2561 and Dow Jones is 4.4379. It is obvious that the prediction result of Dow Jones outperforms the other indexes in this exercise as it has the least MAPE figure of 4.4379. This confirms the speculation that strong-form EMH market should get better result in the above models. Shanghai Composite Index and HSI MAPE values are very close suggesting that China and Hong Kong security market are closely related. In general the improvement of accuracy using wavelet function also only happens in GARCH models. The degree of accuracy in GARCH and its wavelet function are poor compared with that of SVR and LSSVM. As explained in our data section, the pre-processing data method in GARCH cannot use windowize method and it is very likely why its result is so poor. The strength of GARCH is its flexible adaptation of the dynamics of volatilities and its case of estimation when compared to other models. It is a return-based model but it might neglect the important intraday information. E.g. when today’s closing price equals to last day’s closing price, the price return will be zero, but the price variation during today might be volatile. [15] explained the model is not able to capture the information. Despite the renowned reputation in GARCH and previous work on the successful application of GARCH with wavelet based kernel to financial time series, our experiment cannot attain the same result. However, the effect of wavelet based kernel is still a major contributing factor in the overall result in GARCH model. Perhaps another type of GARCH model should be employed to achieve a better result. This will be in our future work and not the scope of this paper. In this section, the focus is to compare and identify the fundamental factors that cause the difference in different models and markets. We simply provide the best model for the above exercises based on our findings.

<table>
<thead>
<tr>
<th>Index Average</th>
<th>Sh Composite</th>
<th>Sh Composite</th>
<th>Hang Sang</th>
<th>Hang Sang</th>
<th>Dow Jones</th>
<th>Dow Jones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast Horizon</td>
<td>4 days</td>
<td>20 days</td>
<td>4 days</td>
<td>20 days</td>
<td>4 days</td>
<td>20 days</td>
</tr>
<tr>
<td>SVR</td>
<td>1.3755</td>
<td>6.4903</td>
<td>2.8785</td>
<td>4.1385</td>
<td>2.0899</td>
<td>2.5235</td>
</tr>
<tr>
<td>WL_sym_svm</td>
<td>1.6965</td>
<td>22.7968</td>
<td>4.5380</td>
<td>8.9549</td>
<td>4.0757</td>
<td>6.5828</td>
</tr>
<tr>
<td>LSSVM</td>
<td>2.0918</td>
<td>3.9494</td>
<td>2.4693</td>
<td>2.7868</td>
<td>1.9144</td>
<td>3.9494</td>
</tr>
<tr>
<td>WL_db_lssvm</td>
<td>2.7785</td>
<td>7.7177</td>
<td>3.6038</td>
<td>5.3428</td>
<td>2.3008</td>
<td>7.7177</td>
</tr>
</tbody>
</table>

Table 2. Various markets performance
Table 3. Various markets performance max and min difference

<table>
<thead>
<tr>
<th>Index</th>
<th>Sh Composite</th>
<th>Sh Composite</th>
<th>Hang Sang</th>
<th>Hang Sang</th>
<th>Dow Jones</th>
<th>Dow Jones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast Horizon</td>
<td>4 days</td>
<td>20 days</td>
<td>4 days</td>
<td>20 days</td>
<td>4 days</td>
<td>20 days</td>
</tr>
<tr>
<td>WL_db_svm</td>
<td>6.2279</td>
<td>95.1529</td>
<td>15.7930</td>
<td>44.2048</td>
<td>14.7343</td>
<td>28.0634</td>
</tr>
<tr>
<td>WL_sym_svm</td>
<td>8.2657</td>
<td>91.0760</td>
<td>23.3956</td>
<td>42.9311</td>
<td>15.5045</td>
<td>38.8243</td>
</tr>
<tr>
<td>WL_sym_lssvm</td>
<td>8.5131</td>
<td>17.5376</td>
<td>18.0802</td>
<td>21.6681</td>
<td>15.0719</td>
<td>17.5379</td>
</tr>
<tr>
<td>Garch</td>
<td>16.0702</td>
<td>51.9487</td>
<td>27.1573</td>
<td>69.7003</td>
<td>20.3877</td>
<td>74.6821</td>
</tr>
<tr>
<td>WL_db_garch</td>
<td>19.6848</td>
<td>96.6332</td>
<td>26.0193</td>
<td>66.3760</td>
<td>18.0263</td>
<td>39.0890</td>
</tr>
<tr>
<td>WL_sym_garch</td>
<td>6.7354</td>
<td>81.8148</td>
<td>13.7217</td>
<td>56.9161</td>
<td>14.9553</td>
<td>32.6312</td>
</tr>
</tbody>
</table>

Table 4. Descriptive statistics for various stock indexes during 2007 to 2009

<table>
<thead>
<tr>
<th>Returns</th>
<th>SH Composite Index</th>
<th>Hang Sang Index</th>
<th>Dow Jones Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistics p-value</td>
<td>h-value</td>
<td>Statistics p-value</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.0567</td>
<td>-0.0393</td>
<td>-0.1006</td>
</tr>
<tr>
<td>Variance</td>
<td>6.1035</td>
<td>7.8655</td>
<td>4.2002</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.0332</td>
<td>0.1709</td>
<td>0.1807</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.1061</td>
<td>6.1697</td>
<td>7.1703</td>
</tr>
<tr>
<td>Normality</td>
<td>24.9141</td>
<td>0.1</td>
<td>206.2428</td>
</tr>
<tr>
<td>Q(6)</td>
<td>6.0892</td>
<td>0.4133</td>
<td>5.427</td>
</tr>
<tr>
<td>Q(6)*</td>
<td>13.2112</td>
<td>0.0398</td>
<td>191.5078</td>
</tr>
<tr>
<td>ARCH(6)</td>
<td>11.7167</td>
<td>0.0686</td>
<td>96.4186</td>
</tr>
</tbody>
</table>

Table 5. Descriptive statistics for various stock indexes during 2003 to 2009

<table>
<thead>
<tr>
<th>Returns</th>
<th>SH Composite Index</th>
<th>Hang Sang Index</th>
<th>Dow Jones Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistics p-value</td>
<td>h-value</td>
<td>Statistics p-value</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.0452</td>
<td>0.0385</td>
<td>-0.0065</td>
</tr>
<tr>
<td>Variance</td>
<td>3.536</td>
<td>3.2458</td>
<td>1.7012</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.2169</td>
<td>0.0918</td>
<td>0.0575</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>5.999</td>
<td>12.3643</td>
<td>14.7956</td>
</tr>
<tr>
<td>Normality</td>
<td>553.6119</td>
<td>0.1</td>
<td>5289</td>
</tr>
<tr>
<td>Q(6)</td>
<td>19.0444</td>
<td>0.0041</td>
<td>9.8543</td>
</tr>
<tr>
<td>Q(6)*</td>
<td>128.4139</td>
<td>0.1</td>
<td>852.7444</td>
</tr>
<tr>
<td>ARCH(6)</td>
<td>83.7537</td>
<td>0.1</td>
<td>366.6877</td>
</tr>
</tbody>
</table>

Notes: Normality is the Bera-Jarque(1981) normality test; Q(6) is the Ljung-Box Q test at 6 order for Raw returns; Q(6)* is LB Q test for squared returns; ARCH(6) is Engle’s (1982) LM test for ARCH effect.
Table 3 shows the difference between maximum and minimum MAPE of the 60 results. This is crucial when selecting which model to use in forecasting. Remember these results are from the extreme volatile period caused by financial tsunami. Combining Tables II and III, Shanghai composite index in SVR model ends up having the best average 1.3755 and the least difference 7.3422 in the 4-day forecast. It is very likely that China stock market is still a close market and the impact of financial tsunami is small. In HSI experiment, LSSVM model has the best average 2.4693 and least difference 7.6761 for 4 days and best average 2.7868 and least difference 9.7202. It should be noted that SVR has the best average 2.8785 and least difference 21.174 for 4 days and best average 4.1385 and least difference 32.1048 which is second to LSSVM in terms of accuracy. As far as the objective of this paper is concerned, we need to find out which is the best model for HSI forecast. From Tables II and III, it is obvious the choice is LSSVM but Table I points to SVR. As Table I is from the most current data while Tables II and III are not, our final recommendation is SVR despite a bigger difference value but it has the smallest MAPE 0.4037. The difference value is a test of the model robustness and the criterion here is having a reasonable value. For the second choice, LSSVM is a good candidate for financial advisor for their decision making.

Tables 4 and 5 report the summary of the descriptive statistics for various stock indexes during the two periods based on log-return analysis. If skewness is negative, it shifts to the left and vice versa. If it is a normal distribution, kurtosis is 3. When kurtosis is greater than 3, it is more outlier-prone than non-normal distribution and vice versa. When normality $h = 1$, it is a normal distribution. When $Q(6) h = 1$, the statistic of raw returns indicates significant autocorrelation. When $Q(6)$* $h = 1$, the statistic of squared raw returns indicates significant correlation. When ARCH(6) $h = 1$, ARCH effect shows significant evidence in support of GARCH effects (i.e. heteroscedasticity). Except 2007 to 2009 Shanghai composite series, others are typically characterized by excessive kurtosis and asymmetry. It can be concluded that the above series are characterized by heteroscedasticity and time-varying autocorrelation; therefore, GARCH class models should fit for forecasting. As seen from Figure 1, Figure 2, Table 3 and Table 4, all series exhibit more variability, skewness, kurtosis and volatility clustering such that nonlinear asymmetric EGARCH model should fit it more accurately. In Table II, all values in GARCH model are from EGARCH model with parameters, R,1,M,1,P,1,Q,2. The result consistent with the statistics findings.

4. Conclusion and Future Work

Based on EMH, the above models have been tested in 3 markets. The winner is SVR model as it produces the best MAPE for the HSI value and can perform equally well in the 3 markets. The accuracy for a long term forecast 20-day or one month is always difficult but the results have demonstrated that it is still possible to get MAPE under 2. It is a significant improvement and very useful tool in financial time series analysis. Decision makers can rely on our models to analyse the market trend or benchmark for investment portfolio. As in the experiment, it is a tedious task to search for the right parameters for the models and so far there is no simple solution to the above problem. The science of forecasting is still relying on trial and error approach. However, the experiments have provided a consistent approach which is to search for the parameters as explained in the above sections using the recent historical data. The disadvantage could be time consuming but it seems the ends justify the means if the objective is achieved.

The consistent performance of the Least Square Support Vector Forecasting model has been demonstrated in experiments especially from Table 2. The above approaches are limited to three forecasting techniques which are GARCH, SVR and LSSVM. In order to increase the predictability of the SVR model, chart pattern is another approach which will be explored. In addition, the chaotic factors of the above markets have not been scrutinized. It will be included in future work in these models. For the time being, it is believed that the above models are useful for handling the current market demand even under extreme condition such as financial tsunami.

Acknowledgements

The authors would like to acknowledge the partial support of the CRG grant G-YM07 of The Hong Kong Polytechnic University.

REFERENCES

June, 2007.


[27] A. Wong. Winner of Bull and Bear, China People’s University Publisher, 2010


