Int. J. Advance Soft Compu. Appl, Vol. 8, No. 3, December 2016
ISSN 2074-8523

Exchange Rate Forecasting Using Modified Empirical Mode Decomposition and Least Squares Support Vector Machine

Nur Izzati Abdul Rashid[1], Ruhaidah Samsudin[2], and Ani Shabri[3]

[1, 2] Fakulti Komputeran,
Universiti Teknologi Malaysia, 81310, Johor, Malaysia
[3] Fakulti Sains, Universiti Teknologi Malaysia, 81310, Johor, Malaysia
e-mail: nur_izzati1812@yahoo.com, ruhaidah@utm.my,
ani@utm.my

Abstract

Forecasting exchange rate requires a model that can capture the non-stationary and non-linearity of the exchange rate data. In this paper, empirical mode decomposition (EMD) is combines with least squares support vector machine (LSSVM) model in order to forecast daily USD/TWD exchange rate. EMD is used to decompose exchange rate data behaviors which are non-linear and non-stationary. LSSVM has been successfully used in non-linear regression estimation problems and pattern recognition. However, its input number selection is not based on any theories or techniques. In this proposed model, the exchange rate is decompose first by using EMD into several simple intrinsic mode oscillations called intrinsic mode function (IMF) and a residual. Permutation distribution clustering (PDC) is used to cluster the IMF and the residual into few groups according to their similarities in order to improve the LSSVM input. After that, LSSVM is used to forecast each of the groups and all the forecasted value is sum up in order to obtain the final exchange rate forecasting value where the best number of input for the LSSVM is determine by using partial autocorrelation function (PACF). The result shows that the modified EMD-LSSVM (MEMD-LSSVM) outperforms single LSSVM and hybrid model of EMD-LSSVM.

Keywords: *Exchange rate, forecasting, EMD, LSSVM, PDC.*

1 Introduction

Exchange rate is the conversion of one currency to another currency. Accurate forecast of exchange rate is important as the exchange rate affect the financial market maturation and development of economy. Both the academic communities and economic are much interest to the issues related to the exchange rate forecasting [1]. The appropriate way of exchange rate forecasting determined the success of many fund managers and business and also very useful to make investment in future.

Many previous researchers have introduced various exchange rate forecasting techniques that based on artificial intelligence (AI) method [1]–[7]. Those methods can be categorized into linear or non-linear techniques. Among them, random walk, autoregressive integrated moving average model and artificial neural network are the most common techniques used. Khashei & Bijari [8] and Khashei et al. [9] stated that one of the prominent models that have been widely used in time series forecasting such as exchange rate, economic and stock problems is ARIMA models where a model with the smallest number of parameter can be obtained to fit well data observed pattern. However, ARIMA models need large amount of data. Unfortunately, in real situations, small data sets should be used to forecast over short period of time.

The introduction of support vector machines (SVMs) by Vapnik [10] has become the solution to the non-linear regression estimation problem. It is successfully been used in forecasting time series. Nevertheless, SVMs is very time consuming because it require complicated computational programming technique. Therefore, Least Squares Support Vector Machine (LSSVM) which is the novel SVMs has been introduced [11]. LSSVM is more straight forward in solving linear problem. LSSVM become an effective tool to model and forecast non-linear characteristic of time series data as it is capable to reach the non-linear system with high precision [12]. Years after the introduction of LSSVM, this model has been successfully used to solve forecasting problems in many fields such as stock price [13], stream flow [14] and water demand [12]. However, LSSVM best input number selection is not based on any theory or techniques [15]. The process to obtain the best number of input is done arbitrarily by starting with the smallest number of input until higher number of input. The number of input that gives the smallest error is selected as the best number of input for LSSVM. This manual process is very time consuming as we need to try the number of input one by one until the smallest error is obtained. It is crucial in obtaining the optimal number of input as it will affect the successful of the LSSVM. Research done by Lin et al. [16] using a hybrid EMD and least squares support vector regression (LSSVR) to forecast foreign exchange rate. Nevertheless, this study does not mentioned the technique or method used to select the best number of input for the LSSVR. Therefore, the best number of input in this paper is obtained by using partial autocorrelation function (PACF).

Many hybrid models have been proposed by the previous researchers to ensure more efficient AI and statistical models. In example, Pai & Lin [17] used hybrid ARIMA and SVMs model in forecasting stock price. He et al. [18] make prediction on exchange rate by using a slantlet denoising least squares support vector regression hybrid methodology. Yang & Lin [6] combined EMD and neural network to forecast exchange rate. The results of these studies show that a hybrid model can defeat single models in forecasting time series. Furthermore, recent studies on hybrid models of LSSVM with other suitable techniques or models show that the hybrid model can outperform other single model [19], [20]. Basically, these studies show that a hybrid model is developed to overcome the drawbacks of the single models and ensure that more accurate and effective model can be obtained to forecast time series data.

Motivated by the successfulness of using hybrid model in forecasting time series data especially LSSVM, this study attempt to hybrid LSSVM with the “decomposition-and-ensemble” principle based on EMD where EMD will decomposed the non-stationary and non-linear behavior of the exchange rate data. Empirical mode decomposition (EMD) that has been founded by Huang et al. (1998) is a technique that decompose adaptive time series for non-linear and non-stationary data by using the Hilbert-Huang transform (HHT). EMD is suitable for financial time series to find tendency of fluctuation where forecasting task is simplified into few simple forecasting subtasks [16]. EMD is effective to assist in designing forecasting models for various applications as it can reveal the hidden patterns and trends of time series [22]. This technique has been used in data analysis and forecast financial time series. However, the application of EMD in exchange rate forecasting is very few. There are very few direct application of EMD in forecasting financial market with high frequency data [23].

In tend to improve the input for the LSSVM that resulting from the decomposition of the original exchange rate data via EMD, a clustering technique which is permutation distribution clustering (PDC) is implemented with the goal to reveal an inherent but latent structure of the data set through partitioning the data set into several groups. Bandt & Pompe [24] has introduced PDC to cluster time series that it is a complexity-based approach that interpret its entropy as a univariate time series complexity measurement. Analysis of data set complexity from various fields has been successfully done by using permutation entropy including engineering [25], geology [26] and medicine [27]. Brandmaier [28] stated few advantages of the permutation distribution which are possession of phase invariance, robust to slow drift in signal and invariant to all monotonic transformation of the underlying time series. Other than that, comparison of time series of varying length is allowed due to permutation distribution’s compressed representation. Since the original data has been decomposed into several IMF components and one residual by EMD, clustering would be an effective technique to formalize their similarity and clustering the component based on their similarities by calculating their dissimilarity matrices.

In this paper, EMD and LSSVM are used to forecast USD/TWD exchange rate which can overcome the limitations of non-stationary and non-linearity by considering that the accuracy of proposed model will increase through the decomposition of exchange rate data. The input of the LSSVM will improve by applying PDC to the IMFs and residual resulting from EMD process where PDC will clustering them into few groups based on their similarities. The proposed model which is named as MEMD-LSSVM is compared with single LSSVM and hybrid of EMD-LSSVM. It shows that MEMD-LSSVM outperform the other models by providing more accurate results.

2 Methodology

2.1 Empirical Mode Decomposition (EMD)

N.E. Huang *et al* in 1998 has found empirical mode decomposition that used the Hilbert-Huang transform (HHT) to decompose adaptive time series for non-linear and non-stationary data into several intrinsic mode oscillations called intrinsic mode function (IMF). The original signal's characteristic information at different time scales contain in the IMF. Therefore, this has encouraged this study to use EMD technique to improve the forecasting model's input structure and forecasting performance due to its suitability to exchange rate data nature which are non-linear and non-stationary. IMF should satisfied two conditions [21]. First, the number of the number of extrema and zero-crossings must be either differs at most by one or equal. Second, the envelope's mean value defines by local minima and maxima should be zero at any point. The decomposition of the time series data is according to the following procedure :

i) Identify all the local minima and maxima of the time series data $m(t)$.
ii) Get the lower envelope $m_l(t)$ and upper envelope $m_u(t)$ of the $m(t)$.
iii) Calculate the first mean value $\mu_1(t)$, that is, $\mu_1(t) = (m_l(t) + m_u(t)) / 2$.
iv) Evaluate the difference between the original time series m(t) and the mean time series $\mu_1(t)$. The first IMF $q_1(t)$ is defined as $q_1(t) = m(t) - \mu_1(t)$.
v) Check whether $q_1(t)$ satisfies the two conditions of an IMF property. Steps (i) - (iii) are repeated if the two conditions are not satisfied in order to find the first IMF.
vi) Once the first IMF is obtained, the above steps are repeated in order to find the second IMF until the final time series $e(t)$ which is a residual component is reach that fulfill the termination criteria which suggest stopping the decomposition procedure.

The original time series $m(t)$ can be obtained by the summation of all the IMF components including the one residual component as Eq.(1) as follows :

$$\boldsymbol{m(t)} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{q_i(t)} + \boldsymbol{e_m(t)} \tag{1}$$

2.2 Least Squares Support Vector Machine (LSSVM)

Least squares support vector machine is a novel approach of the support vector machines (SVMs) that was developed by Johan et al. [11]. Through the upper bound of the generalization error minimization, the over-fitting problem can be managed by the SVMs since it is established on the structural risk minimization (SRM) [29]. Nevertheless, it is time consuming when using SVMs because of the usage of quadratic programming during training process. Hence, a more straight forward approach in solving linear problem has been introduced.

Two important aspects that determine the successful of the LSSVM although it is the simplification of the SVMs procedure are choose a kernel function and determine the kernel parameters. A proper selection of LSSVM free parameter must be done through a methodology so that the LSSVM can get regression that robust to the noisy conditions and the user knowledge regarding the effect of the free parameters values in the problem studied [30].

Given a training set x_i, y_i, $i = 1,2,3, \dots, l$. The input data is represented by x_i while the output data is represented by y_i. The regression function defined by LSSVM is :

$$min J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^{l} e_i^2 \tag{2}$$

subject to

$$y_i = w^T \emptyset(x_i) + b + e_i, i = 1,2,3, \dots, l \tag{3}$$

The weight vector is represented by w, the penalty parameter is γ, the approximation error is e_i, the non-linear mapping function is $\emptyset(\ )$ and the bias term is b. Construct the corresponding Lagrange function by

$$L(w, e, x, b) = \int (w, e) - \sum_{i=1}^{l} \alpha_i \, w^T \emptyset(x_i) + b + e_i - y_i \tag{4}$$

The Lagrange multiplier is α_i. Through differentiate partially with respect to b, w , e_i and x_i, the solutions can be produced by using Karush-Kuhn-Tucker (KKT) :

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{l} \alpha_i \emptyset(x_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{l} \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i \gamma e_i \\ \frac{\partial L}{\partial x_i} = 0 \rightarrow w^T \emptyset(x_i) + b + e_i - y_i = 0 \end{cases} \tag{5}$$

The following equations can be obtained after the elimination of e_i and w

$$\begin{bmatrix} b \\ x \end{bmatrix} = \begin{bmatrix} 0 & I_v^T \\ I_v & \Omega + \gamma^{-1} I \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{6}$$

Where $I_v = [1,1,\ldots,1]^T$, $y = [y_1, y_2, \ldots, y_l]^T$ and $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_l]^T$. Mercer condition is applied to matrix Ω with $\Omega_{km} = \emptyset(x_i)^T \emptyset(x_m), k, m = 1,2,3, \ldots, l$. The LSSVM regression is get from

$$y(x) = \sum_{i=1}^{l} x_i K(x, x_i) + b \tag{7}$$

The kernel function is represented by $K(x, x_i)$.

Two parameters of LSSVM are the kernel parameter, σ^2 and the margin parameter, γ. The optimized value of these parameters is needed and can be obtained through the parameter optimization process. It is important to guarantee that a well performed and accurate model can be obtained. These parameters' values will affect the generalization and training capability of LSSVM [31]. Theoretical technique and cross-validation (CV) in grid search are the frequent approaches in the optimization of LSSVM [29]. The potential shortcomings of the trails and error method can be overcome by grid search method [14].

2.3 Permutation Distribution Clustering (PDC)

Briefly, the occurrence probability of certain patterns of the ranks of value in a time series is assigns by permutation distribution [24]. Subsequences of a fixed length, $\boldsymbol{m}$,is resulting from the partitioning of the time series that also called as embedding in $\boldsymbol{m}$ -space. The embedding can be time-delayed with delay *t* in order to calculate the permutation distribution on a coarser time scales. Calculation of the value's rank is done for each of the subsequence for example the observed values sorting's by product. The distinct rank pattern relative frequency is counted to obtain the permutation distribution. This is also called as ordinal pattern. Identification of each possible rank pattern can be made by a permutation of the values 0 to $\boldsymbol{m-1}$.

Given a time series $\boldsymbol{Y = \{y(i)\}_{i=0}^{T}}$, sampled at equal intervals with $\boldsymbol{y(i)} \in \boldsymbol{R}$, the time delayed embedding of this time series into an $\boldsymbol{m}$-dimensional space with time delay *t*, is $\boldsymbol{Y' = [y(i), y(i+t), x(i+2t), \ldots, y(i+(m-1)t)]_{i=0}^{T'}}$ with a total of $\boldsymbol{T' = T-(m-1)t}$ elements. Computation of the permutation of indices from 0 to $\boldsymbol{(m-1)}$ that puts the $\boldsymbol{m}$ values into sorted order can produce the ordinal pattern for an element $\boldsymbol{y'} \in \boldsymbol{Y'}$. The original order relative to two elements $\boldsymbol{y'(i), y'(j), i \neq j}$ will be kept if they have the same value. There are $\boldsymbol{m}!$ distinct ordinal patterns since $\boldsymbol{m}!$ unique permutations of length $\boldsymbol{m}$.

Let $\Pi(\boldsymbol{y})$ be the permutation that an element $\boldsymbol{y} \in \boldsymbol{R^m}$ undergoes when being sorted. The permutation distribution of $\boldsymbol{Y'}$ is

$$p_\pi = \frac{\#\{y\prime \in Y' | II(y') = \pi\}}{T\prime} \quad (8)$$

In the permutation distribution, the time series frequency of unique patterns is simply represented by the distribution and discarded the ordinal patterns' temporal order. The Shannon entropy of the probability distribution P defines the permutation entropy of order $m \geq 2$ as introduced by Band & Pompe [24] :

$$H(P) = -\sum_{\pi \in S_m} p_\pi log p_\pi \quad (9)$$

Where S_m is the set of all m-permutations.

The divergence between probability distributions is commonly measured by the Kullback-Leibler (KL) divergence or also known as relative Shannon entropy in calculating the relative permutation entropy as relative complexity index. Nevertheless, the KL divergence is not a metric since it violates the triangle inequality. Therefore, the permutation distribution of time series is embedded into a metric space by applying the squared Hellinger distance. The Euclidean norm of the difference of the square root vectors of the discrete probability distribution is equal to the squared Hellinger distance, up to scaling. Let two permutation distribution is represented by $P = (p_1, p_2, \dots, p_n)$ and $Q = (q_1, q_2, \dots, q_n)$. The squared Hellinger distance is :

$$D(P, Q) = \frac{1}{\sqrt{2}} \left\| \sqrt{P} - \sqrt{Q} \right\|_2^2 \quad (10)$$

The KL divergence's Taylor approximation's mean can produced the Squared Hellinger distance. Due to bounded between zero and one, non-negative, symmetric and satisfies the triangle inequality, it is a metric. A distance matrix that can be fed to a clustering algorithm of the researcher's choice is form by the pairwise squared Hellinger distance between the set of permutation distribution. The clustering method that has been chosen is sequential agglomerative hierarchical non-overlaping clustering. A cluster will be assign for each time series at first. Construction of cluster through merging cluster based on a dissimilarity measure is iterated until only single top cluster left. Resulting to a hierarchy of clusters. The calculation of the distance between sets of time series is done by the complete linkage method if not stated which is defined as the dissimilarity between two cluster C_i and C_j :

$$d_{complete}(C_i, C_j) = \max_{x \in C_i, y \in C_j} D(x, y) \quad (11)$$

A dendrogram visualized the binary obtained that represent the tree with branch height that shows the distance between clusters.

3 Data

The data set that is used in this study is the daily exchange rate of United States Dollar to New Taiwan Dollar (USD/TWD) starting from July 2005 until December 2009 with a total of 1131 observations. This data is collected from Board of Governors of the Federal System. Fig.1 below shows the graph of the exchange rate over time. The graph shows that the exchange rate data is non-linear as it does not follow a straight line. Besides that, it also shows that it does not have constant value over time that indicates it as non-stationary. In this study, the sample data is partitioned into two parts which are the first part is for training set and the second part is for testing set. The data from 1 July 2005 to 9 February 2009 (905 observations) are used as training data set while the remaining data from 10 February 2009 to 31 December 2009 (226 observations) are used as testing data set. Training set is used for modeling and the testing set is used to evaluate the performance of the models.

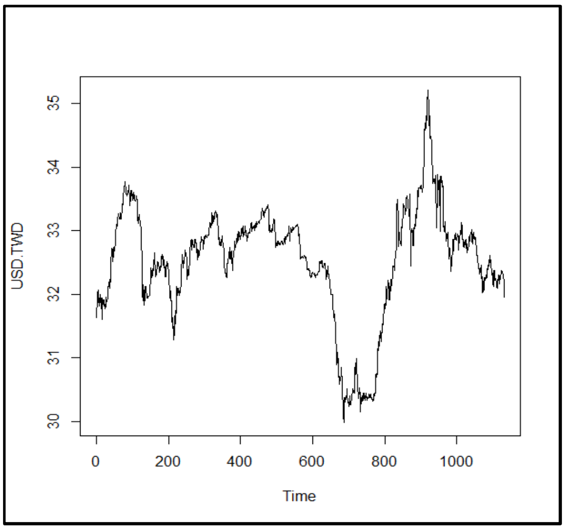

Fig.1: The daily USD/TWD exchange rate graph

4 Proposed MEMD-LSSVM Model

Fig.2 below shows the development of MEMD-LSSVM model in forecasting the exchange rate.

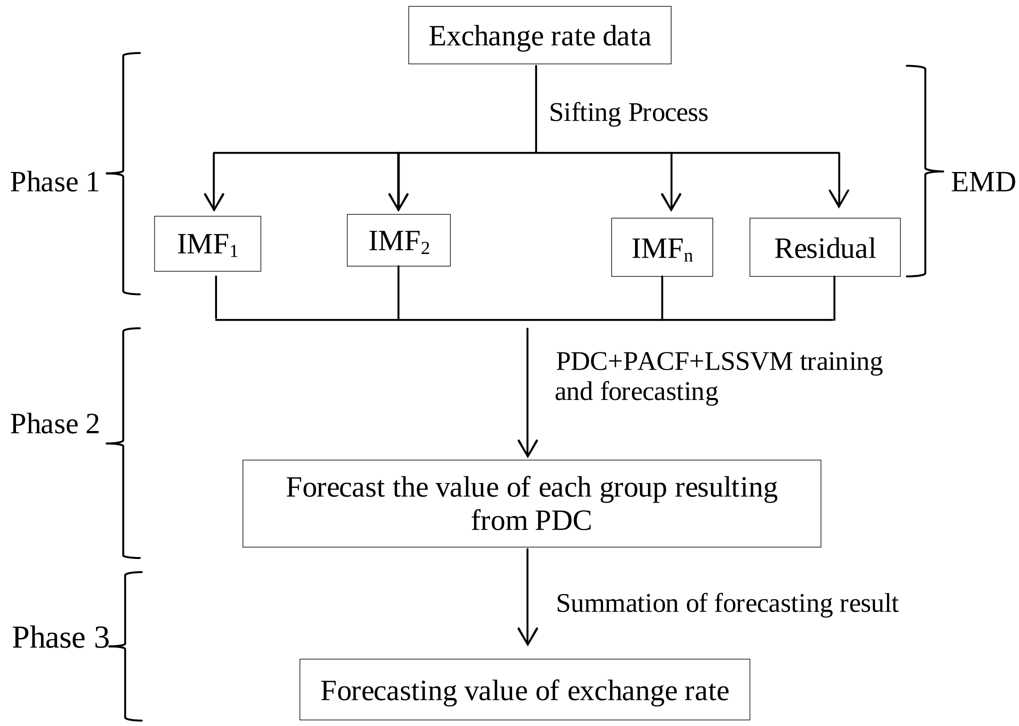

Fig.2: Development Process of EMD-LSSVM Model

Based on the figure above, three phases involve in the development process of EMD-LSSVM. Phase 1 is implementing the EMD technique where the decomposition of the original exchange rate data into several IMF components is occur. n represents the number of IMF components. There are several IMF components and one residual produced.

Training and forecasting each of the IMF components and the residual by using LSSVM occurred during Phase 2. Permutation distribution clustering (PDC) which is a complexity-based approach to cluster time series is implemented before the training process in order to group all the IMF components and the residual into several groups based on their similarities. After that, LSSVM is used to train and forecast each of the groups. The best number of input for the LSSVM is identified by using partial autocorrelation function (PACF) during training process based on the PACF graph of each of the groups.

The forecasting value of each group is sum up together in order to obtain an ensemble of the original exchange rate forecasting value. The actual exchange rate value is compared to the forecasting results. The second trial will start at Phase 2 where the IMF and the residual will be grouped again but this time into different number of groups from the previous trial. Then, the next steps are the same as the first trial where the total value will be trained and forecasted using LSSVM and during Phase 3, sum up the forecasting value of each of the group. There will be few trials that involved the iteration of Phase 2 and Phase 3 until the best forecasting result can be obtained.

4.1 Model Performance Measurement

Mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) that widely used to evaluate the result of time series forecasting are used to measure the performance of the proposed model. These performance evaluations are the error between the actual value and the predicted value. Then, the performance of the proposed model is compared with LSSVM and hybridization of EMD and LSSVM (EMD-LSSVM) to show that the proposed model can forecast more accurate than the other models that do not implement PDC. The model that gives the smallest value of MAE, RMSE and MAPE is the best model. The calculation for these performance measurements are as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2} \quad (12$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - f(x_i)| \quad (13)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - f(x_i)}{y_i}\right| \quad (14)$$

Where the actual value is represented as y_i, the predicted value is represented as $f(x_i)$.

5 Experimental Results and Discussions

In the proposed model, EMD is applied to the exchange rate data in order to decompose the original data into few IMF components and one residual. Fig.3 below shows the decomposition result of the original exchange rate data where the total number of IMF produced are eight and one residual. The IMF components produced starting from the highest to the lowest in term of their frequency.

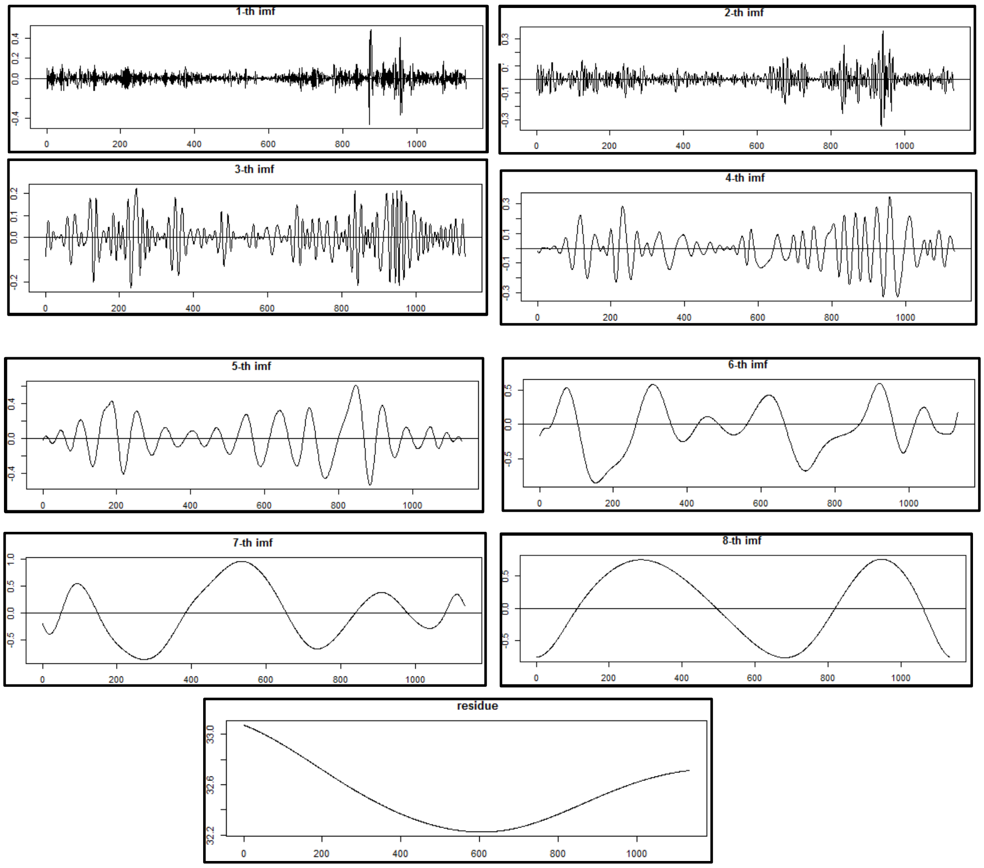

Fig.3: Decomposition of USD/TWD exchange rate via EMD

After the decomposition of the original exchange rate via EMD, PDC is applied for clustering all the IMF components including the residual into several groups based of their similarities. It is one way to improve the input for the LSSVM. In this study, the IMFs and the residual are clustered into 2, 3, 4 and 5. Table 1 shows the list of cluster including the component in group of each cluster.

Table 1 : List of cluster

No. of Cluster	Components
2	G1 = {imf1 - imf2}
	G2 = {imf3 – residual}
3	G1 = {imf1}
	G2 = {imf2}
	G3 = {imf3 – residual}
4	G1 = {imf1}
	G2 = {imf2}
	G3 = {imf3 – imf5}
	G4 = {imf6 – residual}
5	G1 = {imf1}

	G2 = {imf2}
	G3 = {imf3}
	G4 = {imf4 – imf5}
	G5 = {imf6 – residual}

Firstly, all the components are clustered into 2 clusters before the LSSVM forecasting model is build. One of the important steps in forecasting using LSSVM is to determine the best number of input as it will affect the performance of the LSSVM. The best number of input is identified based on the PACF graph by using training data set as studied by Guo et al. [22]. The PACF graph plots the PACF against the lag length. Assume that x_t is the output variable and the PACF graph shows that the partial autocorrelation is out of the 95% confidence interval at lag i, therefore x_{t-i} is one of the input for the LSSVM. Fig.4 below shows PACF of original data and group for 2 cluster. The input variable for the other cluster is shown in Table 2.

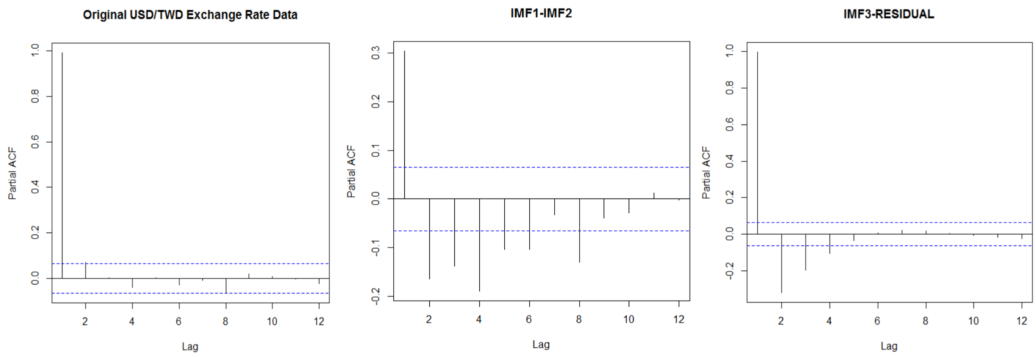

Fig.4: PACF graph of original and group for 2 cluster

Table 2 : Input variables for LSSVM

No. of Cluster	Components	Input Variables
2	G1 = {imf1 - imf2}	{ x_{t-1} - x_{t-6} , x_{t-8}}
	G2 = {imf3 – residual}	{ x_{t-1} - x_{t-4}}
3	G1 = {imf1}	{ x_{t-1} , x_{t-2} , x_{t-4} - x_{t-6}}
	G2 = {imf2}	{ x_{t-1} - x_{t-4} , x_{t-7}}
	G3 = {imf3 – residual}	{ x_{t-1} - x_{t-4}}
4	G1 = {imf1}	{ x_{t-1} , x_{t-2} , x_{t-4} - x_{t-6}}
	G2 = {imf2}	{ x_{t-1} - x_{t-4} , x_{t-7}}
	G3 = {imf3 – imf5}	{ x_{t-1} - x_{t-5} , x_{t-7} , x_{t-8}}
	G4 = {imf6 – residual}	{ x_{t-1} - x_{t-3}}
5	G1 = {imf1}	{ x_{t-1} , x_{t-2} , x_{t-4} - x_{t-6}}
	G2 = {imf2}	{ x_{t-1} - x_{t-4} , x_{t-7}}
	G3 = {imf3}	{ x_{t-1} , x_{t-2} , x_{t-4} - x_{t-8}}
	G4 = {imf4 – imf5}	{ x_{t-1} - x_{t-6} , x_{t-8} - x_{t-12}}
	G5 = {imf6 – residual}	{ x_{t-1} - x_{t-3}}

LSSVM is used to train the data once the best number of input is obtained based on the PACF graph plot. The optimization of LSSVM parameter which are kernel parameter,$\boldsymbol{\sigma^2}$ and the margin parameter, $\boldsymbol{\gamma}$ is needed for each of the group during the training process so that the best parameter for the LSSVM can be obtained. The kernel function that has been chosen is Gaussian radial basis function (RBF) for its good performance than other kernel function. The parameter optimization is important to ensure better performance for the LSSVM. The generalization and training capability of LSSVM is affected by the value of these parameters [31]. Referring to the previous researcher's work, cross validation method and grid search algorithm has been used. 10-fold cross validation was performed to overcome parameter sensitiveness [12]. After the best parameters are obtained, they are used for LSSVM for prediction of the training data set. Lastly, prediction using the testing data set is performed and the forecasting value of each group is summed up in order to get the exchange rate forecasting value. Then, the error between forecasting value and the actual value is calculated according to the performance measurement. The phases starting from implementing PDC to all the IMF components and the residual until forecasting using LSSVM are repeated but by using different number of cluster. Those repetitive phases are continuing processes until the best forecasting result is obtained. The best forecasting result obtained in this study is when the number of cluster is 4.

The comparison between the proposed model with single LSSVM and hybrid model of EMD-LSSVM are made to ensure the capability of the proposed model to perform better in forecasting exchange rate data. The forecasting performance of the proposed MEMD-LSSVM model and the other two models on exchange rate data are shown in Table 3.

Table 3 : Performance of three forecasting models

Models	**MAE**	**MAPE**	**RMSE**
LSSVM	0.094884	0.002858	0.152186
EMD-LSSVM	0.060132	0.001814	0.08771
MEMD-LSSVM	0.059315	0.001789	0.085682

It is clearly shown that MEMD-LSSVM produce better results than LSSVM and EMD-LSSVM models where it gives the smallest error value in term of MAE, RMSE and MAPE. Comparing those three forecasting models, single LSSVM be at the last rank and EMD-LSSVM better than LSSVM. Based on this result, no doubt that the decomposition technique by using EMD is efficient to improve the performance of forecasting model because the non-linear and non-stationary behaviors of the exchange rate data have been decomposed and the hidden patterns can be reveal and understood. The benefit of using EMD is its suitability in decomposing non-stationary and non-linearity of exchange rate data. However, the studies on how to improve the existing forecasting models have never been stop. Therefore, this study tends to apply another approach that can improve the input for the forecasting model through clustering.

The implementation of PDC to the decomposition result of the exchange rate data via EMD make the performance of the forecasting models improved. PDC can reveal inherent but latent structure of the dataset and grouping them. The determination of best input for LSSVM by using PACF promises that only the needed inputs that contribute to the forecasting model is chosen. Therefore, only meaningful inputs to the LSSVM are given. Through specific technique used, we can ensure that the accurate input is chosen to be provided to LSSVM and can avoid human error while doing it manually. As a result, the forecasting model prediction performance will be improved. Besides that, time to identify the best input can be saved instead of manually identify them one by one starting from smallest number of input until higher number of input.

A good forecasting model that can predict the future values accurately is important to ensure that the forecasting result can be used as knowledge, source and information in making plan or decision for the growth of our country's economy. Planning can be made by the economist in order to overcome any possible effect or outcome based on the predicted exchange rate values. This can prevent our country's economy from facing major loss. In case of exchange rate, we can predict tomorrow's value of exchange rate by using today and previous values. The predicted values of exchange rate resulting from the proposed method can be used as a basis to make investment in future. Therefore, our economist and trader can make decision whether they can make profit or they will loss based on the forecasting results and also enhance their trading strategy to manage potential risk of loss. The results can also being used as the evaluation of the risks and benefits attached to the environment of the international business. Other than that, as our country is doing imports and exports, the exchange rate's movement can be very crucial as it will affect the demands for imports and exports. Therefore, the economist can make plan ahead to deal with any possible situation. This indicates that the exchange rate prediction by the proposed method can become useful information in helping the development of our country's economy.

6 Conclusion

A modified EMD-LSSVM model for exchange rate forecasting is developed in this paper. The empirical results shows that with the implementation of decomposition strategy via EMD to the exchange rate data, the non-linear and non-stationary behavior of the exchange rate data can be address effectively and the hidden pattern of the data can be reveal for better understanding resulting in improving the forecasting accuracy. This can be proved with better forecasting result produced by EMD-LSSVM compared to LSSVM. The implementation of PDC in clustering the IMFs and residual resulting from EMD into several groups shows that the proposed MEMD-LSSVM model outperforms LSSVM and EMD-LSSVM. Thus, it can be concluded that PDC give contribution to improve the input for LSSVM in this proposed MEMD-LSSVM

ACKNOWLEDGEMENTS
The authors appreciatively acknowledge the corrections and comments of the anonymous reviewers that significantly improved this paper.

References

[1] L. Liu and W. Wang. 2008. Exchange Rates Forecasting with Least Squares Support Vector Machine, *2008 Int. Conf. Comput. Sci. Softw. Eng.*, vol. 5, no. 4, pp. 1017–1019.

[2] I. a Adetunde. 2011. Forecasting Exchange Rate Between the Ghana Cedi and the Us Dollar Using Time Series Analysis, *African J. Basic Appl. Sci.*, vol. 3, no. 6, pp. 255–264.

[3] J. Kamruzzaman, R. a. Sarker, and I. Ahmad. 2003. SVM based models for predicting foreign currency exchange rates, *Third IEEE Int. Conf. Data Min.*, pp. 557-560.

[4] B. Premanode and C. Toumazou. 2013. Improving prediction of exchange rates using Differential EMD, *Expert Syst. Appl.*, vol. 40, no. 1, pp. 377–384.

[5] G. Sermpinis, C. Stasinakis, K. Theofilatos, and A. Karathanasopoulos. 2015. Modeling, forecasting and trading the EUR exchange rates with hybrid rolling genetic algorithms - Support vector regression forecast combinations, *Eur. J. Oper. Res.*, vol. 247, no. 3, pp. 831–846.

[6] H.-L. Yang and H.-C. Lin. 2011. Applying EMD-based neural network to forecast NTD/USD exchange rate, *7th Int. Conf. Networked Comput. Adv. Inf. Manag.*, no. 64, pp. 352–357.

[7] H.-L. Yang and H.-C. Lin. 2015. Applying the Hybrid Model of EMD, PSR, and ELM to Exchange Rates Forecasting, *Comput. Econ.*, pp.1-18.

[8] M. Khashei and M. Bijari. 2011. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, *Appl. Soft Comput. J.*, vol. 11, no. 2, pp. 2664–2675.

[9] M. Khashei, M. Bijari, and G. A. Raissi Ardali. 2009. Improvement of Auto-Regressive Integrated Moving Average models using Fuzzy logic and Artificial Neural Networks (ANNs), *Neurocomputing*, vol. 72, no. 4, pp. 956–967.

[10]V. Vladimir. 1995. *The Nature of Statistical Learning Theory*, New York:Springer.

[11]S. Johan, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. 2002. *Least Squares Support Vector Machines*, World Scientific.

[12]A. Shabri, R. Samsudin, and U. Teknologi. 2015. Empirical Mode Decomposition – Least Squares Support Vector Machine Based for Water Demand Forecasting, *Int. J. Adv. Soft Comput. Its Appl.*, vol. 7, no. 2.

[13]O. Hegazy, O. S. Soliman, and M. A. Salam. 2014. LSSVM - ABC Algorithm for Stock Price prediction, *Int. J. Comput. Trends Technol.*, vol. 7, no. 2, pp. 81–92.

[14]S. M. Pandhiani and A. Bin Shabri. 2013. Time Series Forecasting Using Wavelet-Least Squares Support Vector Machines and Wavelet Regression Models for Monthly Stream Flow Data, *Open Journal of Statistics*, vol.3, no.3, pp. 183–194. doi: 10.4236/ojs.2013.33021.

[15]S. Ismail, A. Shabri, and R. Samsudin. 2011. A hybrid model of self-organizing maps (SOM) and least square support vector machine (LSSVM) for time-series forecasting, *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10574–10578.

[16]C.-S. Lin, S.-H. Chiu, and T.-Y. Lin. 2012. Empirical mode decomposition–based least squares support vector regression for foreign exchange rate forecasting, *Econ. Model.*, vol. 29, no. 6, pp. 2583–2590.

[17]P.-F. Pai and C.-S. Lin. 2005. A hybrid ARIMA and support vector machines model in stock price forecasting, *Omega*, vol. 33, no. 6, pp. 497–505.

[18]K. He, K. K. Lai, and J. Yen. 2010. A hybrid slantlet denoising least squares support vector regression model for exchange rate prediction, *Procedia Comput. Sci.*, vol. 1, no. 1, pp. 2407–2415.

[19]Z. Mustaffa, Y. Yusof, and S. Kamaruddin. 2014. Social and Gasoline Price Forecasting : An Application of LSSVM with Improved ABC, *Procedia - Soc. Behav. Sci.*, vol. 129, pp. 601–609.

[20]B. Zhu and Y. Wei. 2012. Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology, *Omega (United Kingdom)*, vol. 41, no. 3, pp. 517–524.

[21]N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, N. Yen, C. C. Tung, and H. H. Liu. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Royal Society of London Proceedings Series A*, vol. 454, no. 1971, pp. 903–995.

[22]Z. Guo, W. Zhao, H. Lu, and J. Wang. 2012. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model, *Renew. Energy*, vol. 37, no. 1, pp. 241–249.

[23]L. Hong. 2011. Decomposition and forecast for financial time series with high-frequency based on empirical mode decomposition, *Energy Procedia*, vol. 5, pp. 1333–1340.

[24]C. Bandt and B. Pompe. 2002. Permutation entropy: a natural complexity measure for time series., *Phys. Rev. Lett.*, vol. 88, no. 17, p. 174102.

[25]U. Nair, B. M. Krishna, and V. N. N. Namboothiri. 2010. Permutation entropy based real-time chatter detection using audio signal in turning process, *Int. J. Adv. Manuf. Technol.*, vol. 46, pp. 61–68.

[26]Hao C.Y and T. Q. Zhao. 2010. Regional differentiation based on permutation entropy and its geographical explanation, *Proc. Third Int. Symp. Comput. Sci. Comput. Technol. (ISCSCT 2010)*, no. August, pp. 5–8.

[27]N. Nicolaou and J. Georgiou. 2011. The use of permutation entropy to characterize sleep electroencephalograms, *Clin. EEG Neurosci.*, vol. 42, no. 1, pp. 24–28.

[28]A. M. Brandmaier. 2015. pdc: An R Package for Complexity-Based Clustering of Time Series, *J. Stat. Softw.*, vol. 67, no. 5, pp. 1–23.

[29]M. Afshin, A. Sadeghian, and K. Raahemifar. 2007. On Efficient Tuning of LS-SVM Hyper-Parameters in Short-Term Load Forecasting : A Comparative Study.In *Power Engineering Society General Meeting. IEE*, pp. 1-6.

[30]A. M. Á. Meza, G. D. Santacoloma, C. D. A. Medina, and G. C. Dominguez. 2012. Parameter Selection in Least Squares-Support Vector Machines Regression Oriented , Using Generalized Cross-Validation, *Dyna*, vol. 79, no. 171, pp. 23–30.

[31]S. G. Cao, Y. B. Liu, and Y. P. Wang. 2008. A forecasting and forewarning model for methane hazard in working face of coal mine based on LS-SVM, *J. China Univ. Min. Technol.*, vol. 18, no. 2, pp. 172–176.