Journal of Theoretical and Applied Information Technology
31st January 2017. Vol.95. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

COMPARISON BETWEEN MEMD-LSSVM AND MEMD-ARIMA IN FORECASTING EXCHANGE RATE

[1]NUR IZZATI ABDUL RASHID, [2]ANI SHABRI, [3]RUHAIDAH SAMSUDIN
[1] Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia
[2] Faculty of Science, Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia
[3] Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia
E-mail: [1]nur_izzati1812@yahoo.com, [2]ani@utm.my, [3]ruhaidah@utm.my

ABSTRACT

Due to the non-stationary and non-linearity behaviors of exchange rate data, an appropriate forecasting model that can capture these behaviors is crucial. This paper comparing the performance of modified empirical mode decomposition (EMD) and autoregressive integrated moving average (ARIMA) named as MEMD-ARIMA and modified empirical mode decomposition (EMD) and least squares support vector machine (LSSVM) named as MEMD-LSSVM in forecasting daily USD/TWD exchange rate. EMD technique is firstly used to decompose the exchange rate data that resulting in few intrinsic mode function (IMF) and one residual. In order to improve the result of the EMD so that more effective input can be provided to the forecasting models which are LSSVM and ARIMA, they are clustered into several groups via permutation distribution clustering (PDC). The successfulness of LSSVM in forecasting is depending on the input number selection. The problem is the input number selection is not based on any theories or techniques. Therefore, partial autocorrelation function (PACF) is used in this paper in determining the best number of input for LSSVM. This paper finds that the implementations of PDC has improved the performance of EMD-LSSVM and EMD-ARIMA and also suggest the PDC is suitable either for linear or non-linear model.

Keywords: *Exchange rate, forecasting, Empirical Mode Decomposition (EMD), Least Squares Support Vector Machine (LSSVM), Permutation Distribution Clustering (PDC)*

1. INTRODUCTION

According to [1], issues corresponding to the exchange rate forecasting have attracted economic and academic communities. The conversion of one currency to another currency is referring to the exchange rate. Exchange rate plays an important role on a country's economy as financial maturation and economy's development are affected by exchange rate. Therefore, this has encouraged the research that related to the forecasting model for exchange rate because a better prediction of exchange rate can contribute to the country's economy development. The result from forecasting exchange rate with the proper or suitable techniques become very beneficial in making investment in future and also determine many businesses and fund managers success.

Many previous researchers had used either linear model or non-linear model in forecasting exchange rate data and also hybridized them with other suitable techniques or methods that resulting in a better forecasting results. Therefore, this study tends to apply a technique or method that is suitable and can contribute to both existing linear and non-linear models to perform better and give more accurate forecasting results.

This study tends to hybrid a linear model and a non-linear model with EMD that is "decomposition-and-ensemble" based principle as motivated by pervious researcher's work that prove the successfulness in time series forecasting by using hybrid model. EMD that introduced by N.E. Huang *et.al*[2] is suitable to the behaviors of exchange rate data which are non-stationary and non-linear where this technique use Hilbert-Huang transform (HHT) to decompose adaptive time series for non-stationary and non-linear. Guo *et.al*[3] stated that EMD able to capture trends and hidden patterns of time series. Therefore, it is very useful for financial time series forecasting and data

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

analysis but its application in exchange rate forecasting is very few.

A clustering technique that is a complexity-based approach named as permutation distribution clustering (PDC) has been introduced by Bandt *et.al* [4]. Its entropy is interpreted as a univariate time series complexity measurement. In order to improve the results of EMD that is going to be used as the input for the forecasting models, permutation distribution clustering (PDC) which is a clustering technique is implemented. Through the data set partitioning into several groups, its inherent but latent structure can be revealed. Permutation entropy has been successfully applied in data set complexity analysis from different field such as engineering [5], medicine [6] and geology [7]. There are few advantages of permutation distribution which are robust to slow drift in signal, possession of phase invariance and invariant to all monotonic transformation of the underlying time series. Clustering technique may be a useful technique to formalize the similarities of the IMFs components and the residual resulting from EMD and cluster them through the calculation of their dissimilarities matrices.

Many artificial intelligence (AI) based methods for forecasting exchange rate have been introduced by the previous researchers [1], [8]–[13]. All the AI methods are categorized into two types which are non-linear and linear. Autoregressive integrated moving average (ARIMA) and artificial neural network (ANN) are the most frequent techniques used. The linear forecasting model that has been chosen in this study is ARIMA while for non-linear forecasting model is LSSVM. ARIMA can fit well data observed pattern as it can acquire parameter with the smallest number and in forecasting time series such as stock market and exchange rate, ARIMA is one of the most outstanding models that have been used [14], [15]. However, ARIMA models are a linear model. Thus, ARIMA are restricted to represent non-linearity and non-stationary that present in exchange rate series [16] [17]. It has become one of the limitations of ARIMA because in real world system, the data are often non-linear.

Recent researchers tend to hybrid ARIMA models with other suitable techniques or models in order to overcome the limitation of the ARIMA model and improve the accurateness in forecasting. Due to the difficulties to completely identify the characteristics of time series data in real problem situation, a hybrid model that capable to model non-linear and linear can be a great strategy [14]. Researchers that hybrid ARIMA model in their studies to forecast time series had proven that the hybridization of ARIMA model with the right models or techniques can provide more accurate result compared to the single model [3], [18]–[21]. Therefore, this study tends to use ARIMA model as a linear model that is going to be improved through the hybridization with EMD and PDC.

The introduction of support vector machines (SVMs) by Vapnik [22] has overcome non-linear regression estimation problem. However, SVMs is very time consuming because the computational programming techniques used is complicated. Therefore, the novel SVMs which is least squares support vector machine (LSSVM) that more straight forward has been introduced by Suykens *et.al* [23]. LSSVM has become a strong tool in modeling and forecasting non-linear time series data due to the ability of LSSVM to reach the non-linear system with high precision [16]. After its introduction, LSSVM has been successfully used in several fields such as stock price [24], stream flow [25] and water demand [16] and perform better than other single model such as SVMs and ARIMA. However, the determination of input number of LSSVM is not based on any techniques or theories [26]. The input number is determine manually by trying from smallest number until higher number of input. The number of input that give the smallest forecasting result in term of RMSE, MAPE and MAE is selected as the best number of input. This repetitive process will consume so much time. The successful of LSSVM in forecasting is affected by the optimal number of input. Therefore, this paper use partial autocorrelation function (PACF) graph plot in determining the best number of input.

Research done by Lin *et.al* [27] using the hybridization of EMD and least squares support vector regression (LSSVR) to forecast foreign exchange rates and compare the results with the EMD-ARIMA models. The results show that EMD-LSSVR outperforms EMD-ARIMA. However, the implementation of EMD has improved the accuracy of both LSSVR and ARIMA in forecasting exchange rate. The moving trend of financial time series data is detected by using EMD. The research suggests that the ability of a forecasting model may enhance through time series decomposition.

In this study, it is expected that by clustering the results of EMD via PDC where the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

components are clustered according to their similarities, it will provide more meaningful and better input for the forecasting models. Hence, more accurate forecasting models will be produced.

In this paper, MEMD-LSSVM and MEMD-ARIMA are used to forecast daily USD/TWD exchange rate which can overcome non-stationary and non-linear behaviors of the data by considering that the decomposition of the data will increase the accuracy of the models. The implementation of PDC to the IMF components and the residual will improve the input of the LSSVM. The performance of MEMD-LSSVM and MEMD-ARIMA are compared with LSSVM, EMD-LSSVM, ARIMA and EMD-ARIMA. The successfulness of using PDC to the results of EMD may suggest that it also suitable to be applied to the results of the other time series decomposition technique used to the exchange rate data in order to improve them as the input for the forecasting model.

This study is limited to univariate data where only daily exchange rate is used.

2. METHODOLOGY

2.1 Empirical Mode Decomposition (EMD)

Empirical mode decomposition (EMD) is a technique that decomposes non-stationary and non-linear time series data by using Hilbert-Huang transform (HHT) into few intrinsic mode oscillations named as intrinsic mode function (IMF). All the IMFs hold the behavior of the original signal at different time scales. N.E. Huang *et.al*[2] state two conditions that must be satisfied by the IMF which are the number of zero-crossings and extreme values should be differs at most by one or equal and envelope's mean values defines by local minima and maxima should be zero at any point. The time series data decompositions are following below procedure:

i) Find all the local minima and maxima of the time series data $y(t)$.
ii) Take the lower envelope $y_l(t)$ and upper envelope $y_u(t)$ of $y(t)$.
iii) Compute the first mean value $\mu_1(t)$, that is, $\mu_1(t) = (y_l(t) + y_u(t))/2$.
iv) Calculate the difference between the original time series $y(t)$ and the mean time series $\mu_1(t)$. The first IMF $q_1(t)$ is defined as $q_1(t) = y(t) - \mu_1(t)$.
v) Evaluate whether $q_1(t)$ fulfill the two conditions of an IMF property. If not, repeat steps (i) - (iii) in order to find the first IMF.
vi) After the first IMF is obtained, the above steps are repeated 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 the decomposition procedure to stop.

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

$$\boldsymbol{y(t) = \sum_{i=1}^{n} q_i(t) + e_y(t)} \qquad (1)$$

2.2 Permutation Distribution Clustering (PDC)

Bandt *et.al* [4] stated that permutation distribution will assign the probability of certain patterns of value's rank to occur in a time series. The partitioning of time series will resulting in subsequences of a fixed length, m or also called as embedding in m –space that can be time-delayed with delay t in order to compute the permutation distribution on a coarser time scales. Each of the subsequence's rank of value is calculated in example the observed values sorting's by product. The distinct rank pattern's relative frequency is counted in order to obtain permutation distribution. The possible rank pattern is determine through permutation of the values 0 to $m-1$.

Time series $\boldsymbol{A} = \{\boldsymbol{a(i)}\}_{i=0}^{T}$ that sampled at equal intervals with $\boldsymbol{a(i)} \in \boldsymbol{R}$ is given. $\boldsymbol{A' = [a(i), a(i+t), a(i+2t), \dots, a(i+(m-1)t)]_{i=0}^{T'}}$ is its embedding time-delayed into $\boldsymbol{m}$-dimensional with a total of $\boldsymbol{T' = T-(m-1)t}$ elements. In order to get the ordinal pattern for an element $\boldsymbol{a' \in A'}$, calculate the permutation of indices from 0 to $\boldsymbol{(m-1)}$ that sort the $\boldsymbol{m}$ values in order. Two elements $\boldsymbol{a'(i), a'(j), i \neq j}$ original order relative to each other is kept if they have equal value. Since there are $\boldsymbol{m}$! Unique permutations of length $\boldsymbol{m}$, hence, the distinct ordinal patterns is $\boldsymbol{m}$!.

Let the permutation that an element $\boldsymbol{a \in R^m}$ undergoes when undergoes sorting as $\Pi(\boldsymbol{a})$. The permutation distribution of $\boldsymbol{A'}$ is

$$p_\pi = \frac{\#\{a' \in A' \mid \Pi(a') = \pi\}}{T'} \qquad (2)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Discarded of ordinal patterns' temporal order happen in the permutation distribution and unique patterns in the time series is represented by the distribution.

The Shannon entropy of the probability distribution P defines the permutation entropy of order $m \geq 2$ introduced by Bandt *et.al* [4] :

$$H(P) = -\sum_{\pi \in S_m} p_\pi log p_\pi \quad (3)$$

Where S_m is the set of all m-permutations.

Relative Shannon entropy or also being addressed as Kullback-Leibler (KL) divergence is usually used to measure the divergence between probability distribution. It is also use in representing the entropy's natural expansion as a complex index to compute relative permutation entropy as a relative complexity index. However, it is not a metric because it fails to comply the triangle inequality. Thus, squared Hellinger distance is being used for the embedding of the permutation distribution of a time series into a metric space.

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 :

$$D(P,Q) = \frac{1}{\sqrt{2}} \left\| \sqrt{P} - \sqrt{Q} \right\|_2^2 \quad (4)$$

A distance matrix is produce from the pairwise squared Hellinger distances between the set of permutation distribution. It can become the input of a clustering algorithm of a researcher's choice. A cluster is assigned for each time series at first. Through the iteration of clusters merging that based on a dissimilarity measure, clusters are constructed. This will bring 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. It is defined as the dissimilarity between two cluster C_i and C_j :

$$d_{complete}(C_i, C_j) = \max_{b \in C_i, a \in C_j} D(b,a) \quad (5)$$

A dendogram is usually used to visualize the resulting binary tree by illustrating it with branch heights that shows the distance between clusters.

2.3 Least Squares Support Vector Machine (LSSVM)

Suykens *et.al*[23] have developed the novel approach of support vector machines (SVMs) which is least squares support vector machine (LSSVM). The approach of LSSVM is more straight forward solution to linear problem since SVMs that use quadratic programming during training process is consuming time. Determination of the kernel parameters and choosing a kernel function are the important aspects to ensure the successfulness of LSSVM. A methodology is needed for proper selection of LSSVM free parameter to obtain robustness of regression against the user knowledge about the free parameters values effect in the problem studied and noisy conditions [28].

Given a training set p_i, q_i, $i = 1,2,3, \dots, l$. p_i represent the input data and q_i represent the output data. The definition for regression function by LSSVM is :

$$min J(w,e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^{l} e_i^2 \quad (6)$$

subject to

$$q_i = w^T \emptyset(p_i) + b + e_i, i = 1,2,3, \dots, l \quad (7)$$

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

$$L(w,e,p,b) = \int(w,e) - \sum_{i=1}^{l} \alpha_i \, w^T \emptyset(p_i) + b + e_i - q_i \quad (8)$$

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

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{l} \alpha_i \emptyset(p_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{l} \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i \gamma e_i \\ \frac{\partial L}{\partial p_i} = 0 \rightarrow w^T \emptyset(p_i) + b + e_i - q_i = 0 \end{cases} \quad (9)$$

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

$$\begin{bmatrix} b \\ p \end{bmatrix} = \begin{bmatrix} 0 & I_v^T \\ I_v & \Omega + \gamma^{-1} I \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ q \end{bmatrix} \quad (10)$$

Journal of Theoretical and Applied Information Technology
31st January 2017. Vol.95. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Where $I_v = [1,1,\ldots,1]^T$, $p = [p_1, p_2, \ldots, p_l]^T$ and $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_l]^T$. Mercer condition is applied to matrix Ω with $\Omega_{km} = \emptyset(q)^T \emptyset(q), k, m = 1,2,3, \ldots, l$. The LSSVM regression is get from

$$q(p) = \sum_{i=1}^{l} p_i K(p, p_i) + b \quad (11)$$

The kernel function is represented by $K(p, p_i)$.

The optimize value of LSSVM kernel which are kernel parameter, σ^2 and the margin parameter, γ are needed to promise more accurate and well perform model. Parameter optimization process can determine the optimize values for these parameters. Cao *et.al*[29] stated that LSSVM training and generalization capability are affected by the values of these parameters. Cross-validation and grid search are most commonly used approach for LSSVM parameter optimization and can overcome potential shortcoming of the trails and error method [25], [30].

2.4 Autoregressive Integrated Moving Average (ARIMA)

Autoregressive integrated moving average (ARIMA) or Box-Jenkins (BJ) methodology is a linear model that has the capability to represent either stationary or non-stationary time series. Basically, BJ methodology consist of few steps which are model identification, parameter estimation and diagnostic checking [14], [31]. Once an adequate model is chosen, the model then used to forecast the future value of time series data. The general ARIMA model is known as ARIMA (p,d,q) where p is the order of the autoregressive (AR) part, d is the number of difference and q is the order of moving average (MA) part. Below shows the procedure for BJ methodology :

i) Identify whether the data is stationary or non-stationary. If non-stationary, first or second difference is needed to achieve stationary.
ii) Once the data is stationary, identify the possible model based on the autocorrelation function (ACF) for MA (q) and partial autocorrelation (PACF) for AR (p).
iii) Identify the coefficients and estimate the parameter of the ARIMA model.
iv) Perform diagnostic checking to check for model adequacy. Use PACFs and ACFs of residuals and test statistics to verify whether the model is valid or not. If not, repeat step (ii) to (iv). If valid, use the model for forecasting.
v) Forecast the time series data by using the valid model.

Usually, there are more than one possible ARIMA models that adequate. In this case, Akaike Information Criterion (AIC) value is used to choose the best model where model with the smallest value of AIC is chosen.

3. DATA

Daily exchange rate data of United States Dollar to New Taiwan Dollar (USD/TWD) starting from July 2005 until December 2009 has been used as the data set in this study. The data is collected from Board of Governors of the Federal System website with the total of 1131 observations. The graph of the exchange rate over time is shown in Figure1 below. Based on the graph, it shows the non-linear and non-stationary behaviors of the exchange rate data where it does not have constant value over time that indicates it as non-stationary. The data is partitioned into 80% and 20% in this study. The first set that is used for training process contain the exchange rate data from 1 July 2005 to 9 February 2009 (905 observations) while the second set that is used for testing process contain the exchange rate data from 10 February 2009 to 31 December 2009 (226 observations). During modeling, training data set is used while for the evaluation of the model's performance, testing data set is used.

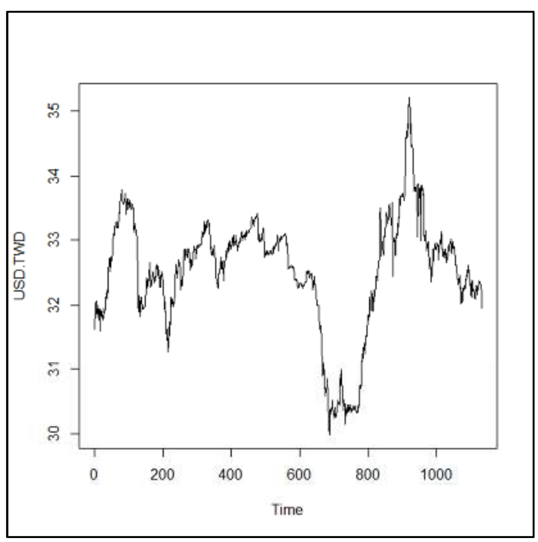


Figure 1 The Daily USD/TWD Exchange Rate Graph

Journal of Theoretical and Applied Information Technology
31st January 2017. Vol.95. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.0 APPLICATION TO DATA

Figure 2 below shows the development of the forecasting model in forecasting the exchange rate.

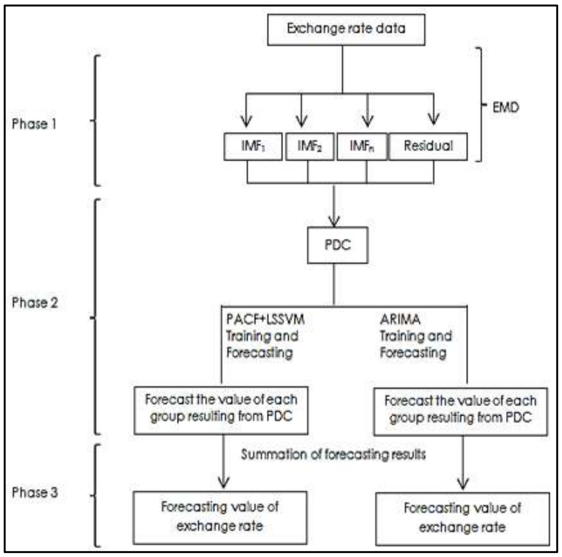


Figure 2 Development Process Of Forecasting Model

This development process is used for the hybrid model of modified EMD and LSSVM (MEMD-LSSVM) and EMD and ARIMA (MEMD-ARIMA). The difference happens during Phase 2 where two different forecasting methods are used. The other steps are the same for both MEMD-LSSVM and MEMD-ARIMA model. There are three phases that involve in the forecasting model development process. EMD technique is implemented during Phase 1 where the original exchange rate data is decomposed into few IMF components and one residual. The number of IMF components is represented by n.

The IMF components and the residual is trained and forecasted by using LSSVM and ARIMA during Phase 2. At the beginning of the training process, PDC is implemented to the IMFs and the residual in order to group them into several groups based on their similarities. PDC identify the components similarity according to differences in their permutation distribution as a proxy for differences in their complexity. The frequency of distinct order patterns in an m-embedding of the components are counted in order to obtain the permutation distribution. It is a repetitive process in order to identify the best number of cluster that will resulting in the best forecasting accuracy where the maximum number of clusters is same as the number of IMFs produced by the EMD. It is the possible cluster that can be form. Once the groups are form, the PACF graph of each of the groups are used to determine the best number of input of LSSVM. Then, LSSVM and ARIMA are used to train and forecast the groups.

Phase 3 involving the summation of the forecasting results of each of the groups in order to obtain the actual forecasting value of the original exchange rate data. Then, the comparison of the actual exchange rate value and the forecasting results is compared.

Phase 2 is repeated for the second trial where different number of groups from the previous trial is used to group the IMFs and the residual. The next steps of the first trials until the completion of Phase 3 are repeated in the second trial where the training and forecasting happened until the summation of the forecasting results of each of the groups. There will be few trials with different number of group while implementing PDC and the iteration of Phase 2 and Phase 3 until obtaining the best forecasting results.

4.1 Model Performance Measurement

Three performance measurements have been used in this study to evaluate the performance of the models which are root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). These performance measurements are widely used to evaluate the result of time series forecasting. These performance evaluations are the error between the actual value and the predicted value. The performance of the MEMD-LSSVM and MEMD-ARIMA are compared with LSSVM, ARIMA, EMD-LSSVM and EMD-ARIMA to show that the implementation of PDC to the modified hybrid models can improve the forecasting accuracy. Best model is determine based on the smallest value of RMSE, MAE and MAPE. Those performance measurements are calculated as follows:

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

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

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

Journal of Theoretical and Applied Information Technology
31st January 2017. Vol.95. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

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

5. RESULTS AND DISCUSSION

Firstly, the original USD/TWD data is decompose into several IMFs and a residual through the implementation of EMD. The decomposition results are shown in Figure 3 below where there are eight IMFs and a residual produce. The IMF components produced starting from the highest to the lowest in term of their frequency. The original signal's characteristic information at different time scales contain in the IMF. It presents the trends and hidden patterns of the exchange rate data.

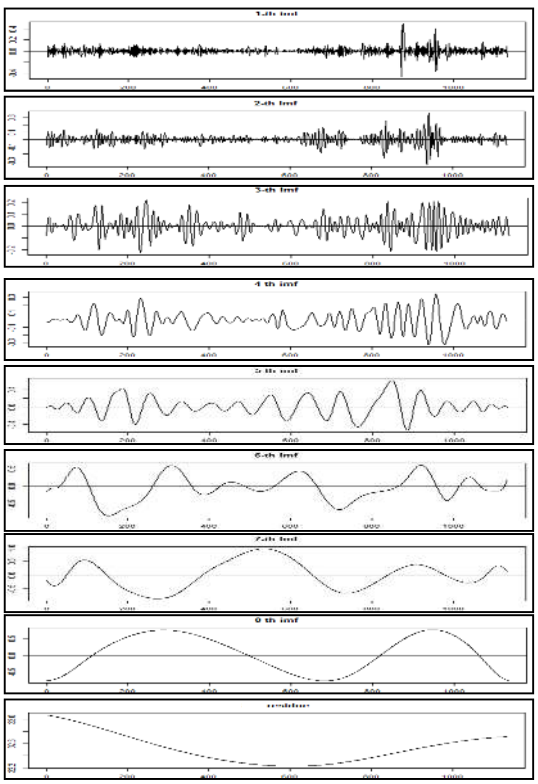


Figure 3 Decomposition of USD/TWD exchange rate via EMD

After that, Phase 2 is begin where all the IMFs and the residual are clustered into few groups based on their similarities by using PDC as a solution in improving the input for the LSSVM. All the IMFs components and the residual in this study are clustered into 2, 3, 4 and 5. The list of clusters and the components in each of their groups are shown in Table 1.

Table 1 List of Clusters

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}

First of all, the IMFs components and the residual are clustered into 2 clusters. Once the groups obtained, the best number of input for each groups is determine for the LSSVM forecasting model based on the PACF graph of the training set that plots the PACF against the lag length. The inputs are specified as the lag variables which their PACF is significant. a_t is assume as the output variable and the partial autocorrelation shown in the PACF graph is out of the 95% confidence interval at lag k. Therefore a_{t-k} is one of the input for the LSSVM. The original data and group for 2 cluster PACF graph are shown in Figure 4 below. Table 2 shows the input variable for 2 and the other cluster.

Journal of Theoretical and Applied Information Technology
31st January 2017. Vol.95. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

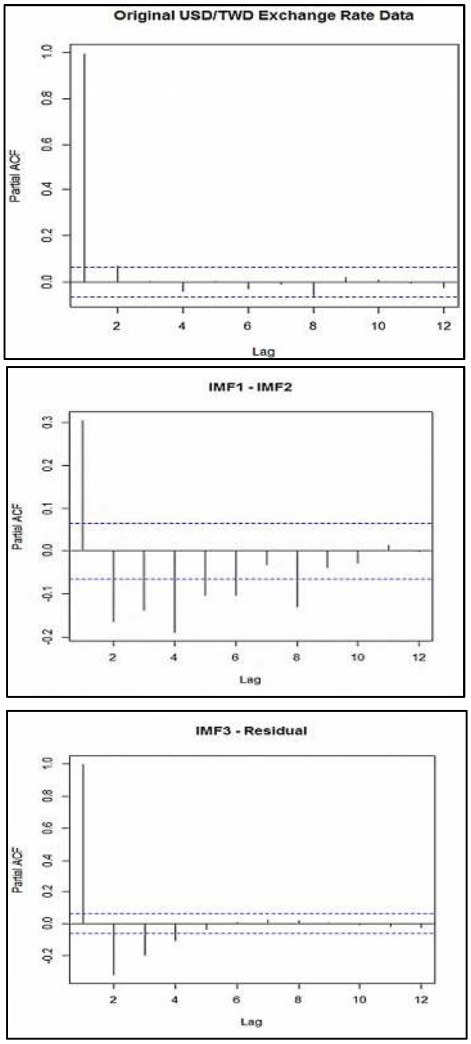


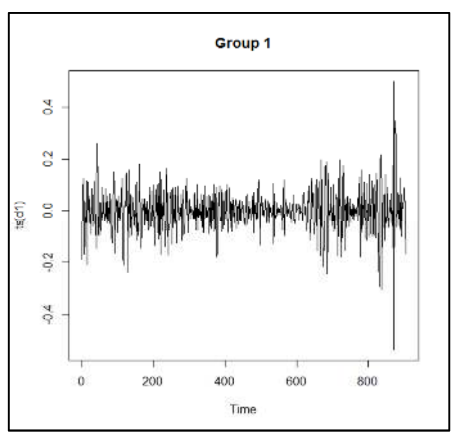
Figure 4 PACF Graph Of Original And Group For 2 Clusters

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}

After the best number of input is obtained, LSSVM is used to train the data. Parameter optimization for kernel parameter, $\boldsymbol{\sigma^2}$ and the margin parameter, $\boldsymbol{\gamma}$ is needed for each groups during the training process in order to obtain the best parameter for LSSVM. Radial basis function (RBF) is used as the kernel function as it performs better than the other kernel function. Based on the previous researcher's work, this study has used cross validation and grid search algorithm. Cao *et.al*[29] stated that parameter sensitiveness is solved by implementing 10-fold cross validation. The best parameters obtained are used for prediction of the training data set by using LSSVM. Then, the prediction of the testing data set is performed and the final forecasting value of the exchange rate data is calculated by the summation of the forecasting value of each of the groups. Lastly, the error between forecasting value and the actual value is calculated according to the performance measurement.

A stationary data is required in forecasting using ARIMA model. Therefore, after the implementation of PDC and the groups of all the components obtained, the data of each group should be identified whether they are stationary or not. Augmented Dickey-Fuller test is used to test the stationarity of the data. Data transformation is needed if the data is non-stationary. The data become stationary after taking first difference. Figure 5 below shows the graph plot of training data of group1 and group2. The graphs that are constant over time indicate that data is stationary. However, the graph of group2 shows that the data is non-stationary. Therefore, first difference is needed.



Journal of Theoretical and Applied Information Technology
31st January 2017. Vol.95. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

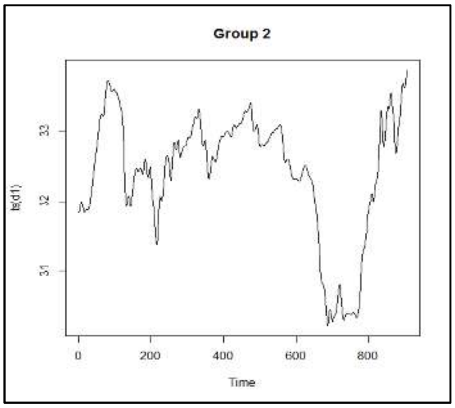


Figure 5 Graph Plot Of Group1 And Group2

Then, the moving average (MA) and autoregressive (AR) model are determine based on the partial autocorrelation function (PACF) and autocorrelation function (ACF) graph. Figure 6 below shows the PACF plot and ACF plot of group1 and group2.

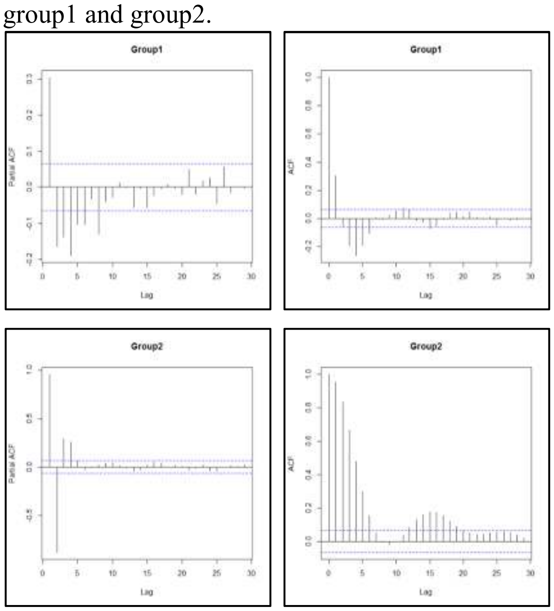


Figure 6 PACF Plot And ACF Plot Of Group1 And Group2

Based on the above graph, few ARIMA models are identified based on the spike lag in the ACF and PACF. The ARIMA model is written as :

ARIMA (p,d,q)

Where p is the MA model, d is the number of difference and q is the AR model.

After the possible ARIMA model is identified, diagnostic checking using Ljung-Box and analysis of ACF and PACF are performed for model adequacy. The aspects that are checked are variance constant, value of mean zero, variance independent and normal distribution. An adequate model is the models that constant in variance, normally distributed with zero mean and independent. The adequate model is chosen as the best model if it has the smallest Akaike Information Criterion (AIC) value. The best model for group1 is ARIMA(4,0,2) while the best model for group2 is ARIMA(3,1,3). These models are then used to forecast the exchange rate data by using the testing data set. Then, the forecasting value of group1 and group2 are summed together to obtain the actual forecasting value. The error between forecasting value and the actual value is calculated according to the performance measurement.

The phases starting from the implementation of the PDC to forecasting the exchange rate are repeated for MEMD-LSSVM and MEMD-ARIMA by using different number of cluster until obtaining the best forecasting result. The number of cluster 4 produce the best forecasting result in this study.

The performance of the MEMD-LSSVM and MEMD-ARIMA are compared with LSSVM, ARIMA, EMD-LSSVM and EMD-ARIMA to show that the implementation of PDC to the modified hybrid models can improve the forecasting accuracy.

The comparison between MEMD-LSSVM and MEMD-ARIMA models with LSSVM, ARIMA, EMD-LSSVM and EMD-ARIMA are made to show that the accuracy of a forecasting model can be improved through the implementation of PDC. Table 3 shows the forecasting performance of MEMD-LSSVM and MEMD-ARIMA and the other four forecasting models. For linear model which is LSSVM, the hybridization of EMD and LSSVM can improve its performance compared to the single LSSVM. Therefore, no doubt that EMD which decomposed the exchange rate data is efficient to improve the performance of a forecasting model because the behaviors of the exchange rate data which are non-stationary and non-linear have been decomposed. Further study in improving the input that results from EMD for a forecasting model has brought to the implementation of clustering technique, PDC to the IMF and residual components. The results shown in Table 3 prove that the clustering technique applied has improved the accuracy of EMD-LSSVM where

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MEMD-LSSVM produces the smallest error in term of MAE, RMSE and MAPE.

Table 3 Performance Of Six 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
ARIMA	0.09122987	0.2753629	0.149555
EMD-ARIMA	0.271958	0.0081	0.390493
MEMD-ARIMA	0.062968	0.001901	0.086667

For non-linear model which is ARIMA, the hybridization of EMD and ARIMA does not perform better than single ARIMA model. This may be due to the difficulties in identification and selection of the ARIMA model for each of the IMF since the original data has been decomposed into different frequency. This problem has been solved by implementing PDC where based on the results in Table 3, the forecasting performance of MEMD-ARIMA is better that EMD-ARIMA and ARIMA. However, MEMD-ARIMA still cannot compete with MEMD-LSSVM that may cause by the problem mentioned before regarding ARIMA model identification. The selection of ARIMA model is depending on the researcher's experience in order to obtain desired forecasting results [27]. Although a linear model is restricted to represent non-linearity and non-stationary that present in exchange rate time series, its hybridization with the suitable techniques or models can overcome those limitations.

Based on the results, the implementation of PDC on the components resulting from EMD has clustered them based on their similarities. This has improved the input for the forecasting models instead of forecasting each of the components individually. The complexity of the input may be reduced. The components that have been clustered were beneficially used as the input to the linear and non-linear forecasting models as they help in improving the performance of the models.

Input is an important factor that ensures the forecasting models can perform optimally and provide better results. A poor forecasting model's accuracy may produce when the input fed to the forecasting model are not good enough or not meaningful to the forecasting models. This has been proven from this study where the improvement made to the input has resulting in better accuracy. The improvement of the input has leads to better performance of both linear and non-linear forecasting models.

6. CONCLUSION

The studies on improving the accuracy of the forecasting models have never been stop. Therefore, this study attempt to apply a clustering approach in order to improve the forecasting model's input through PDC. Based on the results obtained, it shows that the input for the forecasting models have been improved through the implementation of PDC where with PDC, both the linear and non-linear forecasting models produced more accurate results compared to the forecasting model without PDC. This suggests that the implementation of PDC is suitable either for linear or non-linear model and can contribute to more accurate forecasting results. Other than that, the decomposition strategy of EMD to the exchange rate data also contributes in addressing the non-linear and non-stationary behavior of the data.

ACKNOWLEDGEMENT

This research is supported by FRGS grant. The authors acknowledged the Ministry of Higher Education and Universiti Teknologi Malaysia for the research grant. The authors also appreciatively acknowledge the corrections and comments of the anonymous reviewers that significantly improved this paper.

REFERENCES:

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

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

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

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

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

[5] U. Nair, B. M. Krishna, and V. N. N. Namboothiri, "Permutation entropy based real-time chatter detection using audio signal in turning process", *Int. J. Adv. Manuf. Technol.*, vol. 46, no. 1–4, 2010, pp. 61–68.

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

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

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

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

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

[11] G. Sermpinis, C. Stasinakis, K. Theofilatos, and A. Karathanasopoulos, "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, 2015, pp. 831–846.

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

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

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

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

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

[17] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", *Neurocomputing*, vol. 50, 2003, pp. 159–175.

[18] A. J. Conejo, M. a Plazas, R. Espínola, S. Member, and A. B. Molina, "Day-Ahead Electricity Price Forecasting Using the Wavelet Transform and ARIMA Models", *IEEE Trans. Power Syst.*, vol. 20, no. 2, 2005, pp. 1035–1042.

[19] O. Valenzuela, I. Rojas, F. Rojas, H. Pomares, L. J. Herrera, a. Guillen, L. Marquez, and M. Pasadas, "Hybridization of intelligent techniques and ARIMA models for time series prediction", *Fuzzy Sets Syst.*, vol. 159, 2008, pp. 821–845.

[20] D. Omer Faruk, "A hybrid neural network and ARIMA model for water quality time series prediction", *Eng. Appl. Artif. Intell.*, vol. 23, 2010, pp. 586–594.

[21] C. H. Aladag, E. Egrioglu, and C. Kadilar, "Forecasting nonlinear time series with a hybrid methodology", *Appl. Math. Lett.*, vol. 22, no. 9, 2009, pp. 1467–1470.

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

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

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

[25] S. Pandhiani and A. Shabri, "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, 2013, pp. 183–194. doi: 10.4236/ojs.2013.33021.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[26] S. Ismail, A. Shabri, and R. Samsudin, "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, 2011, pp. 10574–10578.

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

[28] L. Li-Xia, Z. Yi-Qi, and X. Y. Liu, "Tax forecasting theory and model based on SVM optimized by PSO", *Expert Syst. Appl.*, vol. 38, no. 1, 2011, pp. 116–120.

[29] S. G. Cao, Y. B. Liu, and Y. P. Wang, "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, 2008, pp. 172–176.

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

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