International Journal of Innovative Computing, Information and Control
Volume **8**, Number **3(A)**, March **2012**

ICIC International ©2012 ISSN 1349-4198
pp. 1917–1930

WAVELET-BASED RELEVANCE VECTOR REGRESSION MODEL COUPLED WITH PHASE SPACE RECONSTRUCTION FOR EXCHANGE RATE FORECASTING

Shian-Chang Huang[1] and Chia-Hsun Hsieh[2]

[1]Department of Business Administration
National Changhua University of Education
No. 2, Shi-Da Road, Changhua 500, Taiwan
shhuang@cc.ncue.edu.tw

[2]Department of Finance
National Central University
No. 300, Jhongda Rd., Jhongli City, Taoyuan County 32001, Taiwan
hcsvincent@hotmail.com

Received December 2010; revised June 2011

Abstract. *Due to the high risk associated with international transactions, exchange rate forecasting is a challenging and important field in modern time series analysis. The difficulty in forecasting arises from the nonlinearity and non-stationarity inherent in exchange rate dynamics. To address these problems, this study proposes a hybrid model that couples two effective feature extraction techniques, phase space reconstruction, and wavelet analysis, with a Relevance Vector Regression (RVR) model to forecast chaotic exchange rates. The time series inputs are first mapped into high-dimension phase space, and then the phase space signal is decomposed on a wavelet basis to analyze its dynamics under various frequencies. Finally, each wavelet component is fed into a local RVR to perform non-parametric regression and forecasting. Compared with other forecasting models, such as support vector machines (SVR), RVR, GJR-GARCH or pure wavelet-base models, the proposed model performs best and statistically improves forecasting performance under root mean square error (RMSE), mean absolute error (MAE) and directional symmetry (DS).*

Keywords: Chaos theory, Phase-space reconstruction, Wavelet analysis, Time series forecasting, Relevance vector regression, Support vector regression

1. **Introduction.** The recent financial tsunami around 2008 to 2010 has caused great loss to investors in currency markets. Owing to the high risk associated with international transactions, exchange rate forecasting is a challenging and important field in modern financial analysis. The difficulties of exchange rate forecasting arise from inherent nonlinearity and nonstationarity in exchange rate dynamics. To address these problems, this study develops a hybrid model that integrates two feature extraction techniques with Relevance Vector Regression (RVR, Tipping [37]) to perform forecasting.

Traditional autoregressive integrated moving average (ARIMA) and GARCH (Bollerslev [4]) models are popular in financial forecasting. However, they are parametric models with strong assumptions that the time series must be stationary. Recent nonlinear nonparametric models – Artificial Neural Networks (ANNs, Kurban and Filik [25]) and fuzzy systems (Abdollahzade et al. [2], Chang et al. [5]) do not suffer from the above limitations. ANN systems are typically classified as "data-driven" methods. An enhanced model from ANNs was recently proposed and found to perform well in time series forecasting, namely Support Vector Machines (SVMs, Vapnik [38]). The SVM model is based on the structural risk minimization (SRM) principle, contrasted to ANNs, which embody the Empirical Risk Minimization (ERM) principle. Under the SRM principle, SVMs solve a convex optimization problem with a unique solution. SVMs typically outperform ANNs and traditional approaches (Perez-Cruz et al. [32], Kim [22], Huang et al. [15], Ince and Trafalis [19], Hong et al. [16], Chen and Jeong [7], Chen and Chen [6], Begum et al. [3], Chen et al.[8]).

However, there are still some disadvantages in the SVR model (see Tipping [37]). First, the number of support vectors grows linearly with increasing training data. Too many support vectors may cause the problem of over-fitting. Second, SVR predictions are not probabilistic. A regression problem needs a probabilistic framework to capture uncertainty in prediction. Third, many SVR parameters, such as the trade-off, insensitivity parameters and kernel width, have to be determined in the initial stage. Fourth, SVM technique is not always able to construct parsimonious models for system identification. Finally, the kernel function must satisfy the Mercer's condition.

The first unique feature of the proposed approach is to replace SVR by RVR (Relevance Vector Regressions, Tipping [37]). RVR overcomes the disadvantages of SVR by embedding SVR into a Bayesian framework. The Bayesian framework determines optimal parameters automatically in the learning process. The RVR significantly reduces the numbers of support vectors, and all the parameters and hyperplanes are estimated in probability. Consequently, RVR can produce sparser models and effectively reduce the risk of overfitting. Recent researches have demonstrated better forecasting capacity of RVRs (Tipping [37], Li et al. [27]). Instead of using a global RVR for prediction, the proposed system employs a group of local RVRs to predict each time-frequency (wavelet) component. Because signal dynamics are complex and heavily dependent on time scales, local prediction on each wavelet component is an effective method to increase forecasting accuracy.

The second unique feature of the proposed approach is that, instead of performing predictions in raw data space, this study constructs forecasting models in the phase space. Prior studies have clearly shown that the behavior of exchange rates can be characterized by dynamic chaotic systems. Scheindman and Lebaron [34], Frank and Stengos [10], Schwartz and Yousefi [35] and Wang et al. [39] found chaotic behavior in financial markets such as the stock market, foreign exchange markets, and the futures market. By phase space (Packard et al. [31]), chaos theory offers a method to capture the attractor trajectory and reveals characteristic information of a chaotic dynamic system. Traditional statistical techniques cannot capture the key information easily. Related works include: Lisi and Schiavo [28], Iseri et al. [18], and Coban and Buyuklu [9].

The third unique feature of the proposed approach is the application of wavelet analysis to extract signal features over multiple time scales to enhance the performance of local predictors. For instance, high-frequency time scales are better for analyzing short-interval features, such as volatility; on the contrary, low-frequency time scales are suitable for recognizing long-interval features, such as trends or long-term patterns. Forecasting models operating in the time domain are difficult to track sudden transients in stock prices. Therefore, the best solution is to transform the financial time series to a domain capable of identifying these key features with compact representation.

Based on the ability of extracting key features in signal processing, wavelet analysis has been applied in financial studies. Kim and In [23] investigated the relationship between stock returns and inflation through wavelet analysis. Yousefi et al. [43] used a wavelet-based model to forecast oil price and investigated whether futures markets are efficiently priced. Mitra [30] applied a wavelet filtering based approach to analyze the relationship among money, output and price for the Indian economy. Related works about the application of wavelet analysis include Karuppiah and Los [20], In and Kim [17], Kim and In [24], Fernandez [11], Gallegati [12], Zhang et al. [45] and Rua and Nunes [33].

The main innovation of this paper lies in combining wavelet analysis, chaos analysis and the RVR model to form a powerful prediction system. In the first stage, this study utilized delay coordinate embedding to draw the trajectory in the high dimensional phase (or state) space. In the second stage, wavelet analysis helps to decompose the chaotic series into several time-frequency components, and each component is fed into a local RVR to perform forecasting. The sum of all local predictions makes up the final predictions. The experimental results suggest that the proposed hybrid model outperforms all other models, including pure models, wavelet-based models and GARCH models (Glosten et al., [13]). The empirical findings clearly reveal that wavelet analysis and the chaos theory significantly improve exchange rate forecasting.

Performing regression in an efficient subspace makes the proposed model more sparse and parsimonious than traditional SVM models, and reduces the risk of overfitting. Similar works related to this study include Zhang et al. [44], Wu and Chang [41], Fernandez [11], Ma and Xu [29] and Yeh et al. [42]. Zhang et al. [44] implemented a wavelet-based prediction system using neural networks. The weakness of their system comes from many drawbacks of NNs. Han et al. [14] and Ma and Xu [29] developed RBF-NN (radial basis function neural network) predictors in phase space. Their systems are more effective than wavelet-based models, but still suffer the drawbacks of NNs. Fernandez [11] compared wavelet-based and SVM-based forecasts, and showed that wavelet-based models outperform SVM-based models. His wavelet model is actually a multi-resolution ARIMA model. The parametric ARIMA model cannot compete with our nonparametric RVR forecaster. More importantly, our wavelet-based model operates in phase space, which is more efficient than pure wavelet-based models. Similar to our work, Wu and Chang [41] and Lau and Wu [26] implemented a phase space support vector regression (SVR) model for time series forecasting. Because system dynamics in phase space are still complex and dependent on time scales, different from their method, this study implements local wavelet-based RVRs in phase space, which overcomes the disadvantages of SVR, and better fits local phase (state) dynamics in each time scale. Yeh et al. [42] constructed a multiple-kernel support vector regressor in time domain for stock price forecasting. Although it outperforms traditional NN and SVM models, it is not parsimonious enough to compete with phase space and wavelet-based forecasters. Its two-stage learning algorithm is complicated, and a large number of model parameters need to be optimized first. Their predictor also suffers from drawbacks of time domain models.

The remainder of this paper is organized as follows. Section 2 introduces the relevance vector regressions. Section 3 introduces phase space reconstructions and the proposed system. Section 4 exhibits the data sets, and reports experimental results measured by RMSE, MAE and DS. The Wilcoxon signed-ranks test (Wilcoxon [40]) is applied to test the performance among each model. Section 5 draws the conclusions.

2. Relevance Vector Regressions.

Given an input-target training pair $S = \{X_t, Y_t\}_{t=1,\ldots,n}$, assuming that targets are independent and the noise of the data is Gaussian with variance σ^2, the probability distribution of Y_t can be written as

$$p(\boldsymbol{Y} \left|\boldsymbol{w}, \sigma^2\right.) = (2\pi\sigma^2)^{N/2} exp\left(-\frac{\left\|\boldsymbol{Y} - \Phi\boldsymbol{w}\right\|^2}{2\sigma^2}\right) \tag{1}$$

where $\boldsymbol{Y} = [Y_1, \ldots, Y_n]^T$, $\boldsymbol{w} = [w_1, \ldots, w_n]^T$, Φ is a matrix in which the rows contain the response of the kernel function to inputs, and the matrix function Φ can be represented as

$$\Phi = \begin{bmatrix} 1 & K(X_1, X_1) & K(X_1, X_2) & \cdots & K(X_1, X_N) \\ 1 & K(X_2, X_1) & K(X_2, X_2) & \cdots & K(X_2, X_N) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & K(X_N, X_1) & K(X_N, X_2) & \cdots & K(X_N, X_N) \end{bmatrix} \tag{2}$$

Rather than attempting to look for optimal weight parameters of the model, a prior distribution is defined over each of the weights. In the RVR framework, Gaussian prior distributions are chosen:

$$p(w_t\,|\alpha_t) = \sqrt{\frac{\alpha_t}{2\pi}} exp\left(-\frac{(w_t \alpha_t^2)}{2}\right) \tag{3}$$

where α_t is the hyperparameter that governs the prior distribution defined over the weight w_t.

Adopting the Markov property, we obtain the following posterior distribution which follows normal distribution:

$$p(Ypred\,|\boldsymbol{Y}) = \int p(Ypred\,\left|\boldsymbol{w}, \boldsymbol{\alpha}, \sigma^2\right.) p(\boldsymbol{w}, \boldsymbol{\alpha}, \sigma^2\,|\boldsymbol{Y}) d\boldsymbol{w} d\boldsymbol{\alpha} d\sigma^2 \tag{4}$$

where *Ypred* is the model prediction. With the prior and the likelihood distributions, the posterior distribution over the weights can be calculated through the Bayes' rule:

$$p(\boldsymbol{w}\,\left|\boldsymbol{Y}, \boldsymbol{\alpha}, \sigma^2\right.) = \frac{p(\boldsymbol{Y}\,|\boldsymbol{w}, \sigma^2) p(\boldsymbol{w}\,|\boldsymbol{\alpha})}{p(\boldsymbol{Y}\,|\boldsymbol{\alpha}, \sigma^2)} \tag{5}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]$. The resulting posterior distribution over the weights is the multivariate Gaussian distribution

$$p(\boldsymbol{w}\,\left|\boldsymbol{Y}, \boldsymbol{\alpha}, \sigma^2\right.) = (2\pi)^{-(N+1)/2} \left|\sum\right|^{-1/2} exp\left(-\frac{(\boldsymbol{w}-\mu)^T \Lambda^{-1} (\boldsymbol{w}-\mu)}{2}\right) \sim N\left(\boldsymbol{\mu}, \sum\right) \tag{6}$$

where the mean and covariance are respectively denoted as

$$\boldsymbol{\mu} = \sigma^{-2} \sum \Phi^{\mathrm{T}} \boldsymbol{Y} \tag{7}$$

$$\sum = (\sigma^{-2} \Phi^T \Phi + \boldsymbol{A})^{-1} \tag{8}$$

with $\boldsymbol{A} = diag[\alpha_0, \ldots, \alpha_N]$. The likelihood distribution over the training targets are computed by integrating the weights and simplified as

$$p(\boldsymbol{Y}\,\left|\boldsymbol{\alpha}, \sigma^2\right.) = (2\pi)^{-N/2} \left|\Lambda\right|^{-1/2} exp\left(-\frac{\boldsymbol{Y} \Lambda^{-1} \boldsymbol{Y}^T}{2}\right) \sim N(0, \Lambda) \tag{9}$$

where the covariance is written by $\Lambda = \sigma^2 \boldsymbol{I} + \Phi \boldsymbol{A}^{-1} \Phi^T$. Given Formula (6) to Formula (9), we rewrite Formula (4) by the following:

$$p(Ypred\,|\boldsymbol{Y}) = \int p(Ypred\,\left|\boldsymbol{w}, \boldsymbol{\alpha}, \sigma^2\right.) p(\boldsymbol{w}\,\left|\boldsymbol{Y}, \boldsymbol{\alpha}, \sigma^2\right.) p(\boldsymbol{\alpha}, \sigma^2\,|\boldsymbol{Y}) d\boldsymbol{w} d\boldsymbol{\alpha} d\sigma^2 \tag{10}$$

In the RVR framework, the estimated value of the weights is obtained by the mean of posterior distribution denoted as Formula (6), which is the Maximum a Posteriori (MAP) estimator. The MAP estimator of the weights determination depends on hyperparameters α and the noise σ^2 in which the two estimates are estimated by maximizing the marginal likelihood

$$(\hat{\boldsymbol{\alpha}}, \hat{\sigma}^2) = \arg\max\, p(\boldsymbol{\alpha}, \sigma^2\,|\boldsymbol{Y}) \tag{11}$$

and the posterior distribution of corresponding output is obtained by the predictive distribution

$$p(Ypred\,|\boldsymbol{Y}) = \int p(Ypred\,\left|\boldsymbol{w}, \hat{\boldsymbol{\alpha}}, \hat{\sigma}^2\right.)p(\boldsymbol{w}\left|\boldsymbol{Y}, \hat{\boldsymbol{\alpha}}, \hat{\sigma}^2\right.)d\boldsymbol{w} \tag{12}$$

The maximization of the marginal likelihood with respect to α and σ^2 is performed iteratively, because there is no closed solution. In practice, during the iterative re-estimation many of the hyperparameters α_t approach infinity, yielding a posterior distribution (5) of the corresponding weight w_t that tends to be a delta function centered around zero. The corresponding weight is thus deleted from the model, as well as its associated basis function $\phi_t(X_t)$. The model is built on the few training examples whose associated hyperparameters do not go to infinity during the training process, leading to a sparse solution. These remaining examples are called the relevance vectors.

3. **Feature Extractions and the Proposed System.**

3.1. **Reconstructed phase space.**

In order to estimate the internal system information from complex time series data, Packard et al. [31] brought up delay-coordinate state space reconstruction to reconstruct chaos attractor. Consider a time series $S_t = (x_1, x_2, \ldots, x_n)$, on the basis of Takens' Embedding Theorem (Takens [36]), the system trajectory that positioned in m-dimension Euclidean space can be reconstructed in $(2m + 1)$-dimension space by delay coordinate. We can define a m-dimension n-delay state vectors as

$$X_t = (x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(m-1)\tau}), \quad t = 1, 2, \ldots, n_m \tag{13}$$

where $n_m = n - (m-1)\tau$, m is the embedding dimensions and τ is a time delays. Given a time series S_t, an m-dimension phase space can be extended

$$PhS = \begin{bmatrix} x_1 & x_{1-\tau} & x_{1-2\tau} & x_{1-3\tau} & \cdots & x_{1-(m-1)\tau} \\ x_2 & x_{2-\tau} & x_{2-2\tau} & x_{2-3\tau} & \cdots & x_{2-(m-1)\tau} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n_m} & x_{n_m-\tau} & x_{n_m-2\tau} & x_{n_m-3\tau} & \cdots & x_{n_m-(m-1)\tau} \end{bmatrix} \tag{14}$$

According to the above definition, the important work is to choose optimal embedding dimension m and delay time τ when reconstruct a suitable phase space.

To reconstruct the phase space, good choices for time lag τ and embedding dimension m are needed. This study uses the first minimum of the mutual information function (Abarbanel [1]) $I(\tau)$ to determine τ:

$$I(\tau) = \sum_{t=1}^{N-\tau} P(x_t, x_{t+\tau}) \log_2 \left(\frac{P(x_t, x_{t+\tau})}{P(x_t)P(x_{t+\tau})} \right), \tag{15}$$

where $P(x_n)$ is the probability density of x_n, while the $P(x_n, x_{n+\tau})$ is the probability density of x_n and $x_{n+\tau}$.

The false nearest neighbor (FNN, Kennel et al. [21]) method is a approach to find the optimal embedding dimension. The idea of the algorithm false nearest is the following. For each point in the time series look for its nearest neighbor in a m-dimensional space. Calculate the distance. Iterate both points and compute

$$R_i = \frac{\|\mathbf{Z}_{i+1} - \mathbf{Z}_{j+1}\|}{\|\mathbf{Z}_i - \mathbf{Z}_j\|} \tag{16}$$

If R_i exceeds a given heuristic threshold R_t, this point is marked as having a false nearest neighbor. The criterion that the embedding dimension is high enough is that the fraction of points for which $R_i > R_t$ is zero, or at least sufficiently small.

3.2. **The proposed system.** The procedure of the new approach is as follows: (1) AMI and FNN are employed to obtain the optimal delay time and embedding dimension for the target index for phase space reconstruction. Based on this technique, this study transforms input variables to phase space which is better for capturing the essential dynamics of a chaotic system. (2) Wavelet analysis is applied to decompose the phase space signal into several time-frequency components for multi-resolution predictions, because the system dynamics differ for each time-frequency (wavelet) component. (3) Each wavelet component is finally fed into a local RVR for prediction. Figure 1 shows the overall flow diagram of the hybrid model.

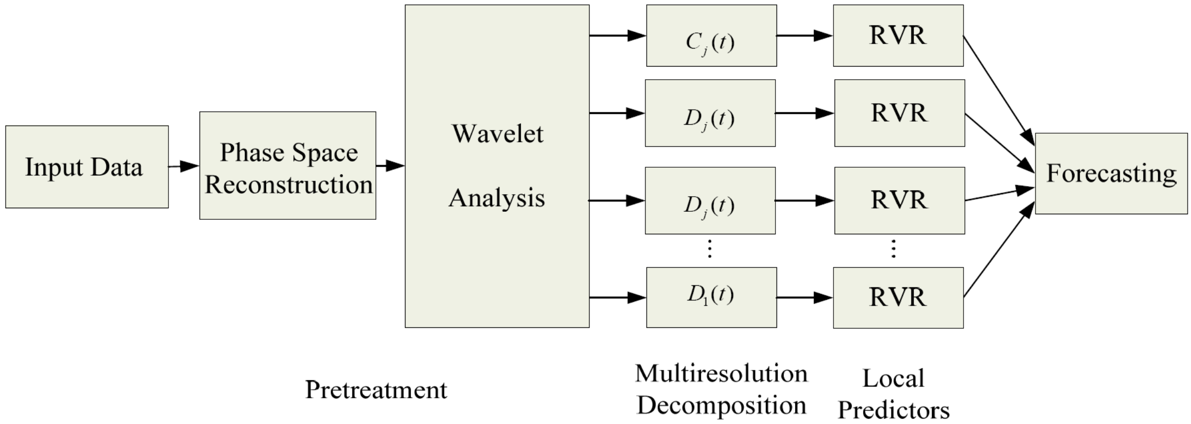

FIGURE 1. Flow diagram of the hybrid model

4. Experimental Analysis and Results.

4.1. **Data description and performance criteria.** This study used monthly data for the experiment running from January 1991 to Jun 2009. All data were sampled from DataStream. The target outputs of this study cover six monthly exchange rate returns, namely GBP/USD, JPY/USD, NTD/USD, DEM/USD, FRF/USD and ITL/USD. The return values were calculated by the following formula:

$$Y_t = \ln(y_t) - \ln(y_{t-1}) = \ln\left(\frac{y_t}{y_{t-1}}\right), \quad t = 2, 3, \ldots, n \tag{17}$$

where Y_t represents the return, y_t represents the raw data. Each data set included 220 observations, divided into training sets and test sets, 160 observations for training sets and 60 observations for test sets. However, exchange rate predictions were difficult due to the complex factors that influence exchange rate movements. Consequently, macroeconomic factors, which may affect exchange rate directly or indirectly, were also selected as input variables. Table 1 lists all the target variables and explanatory variables. Table 2 summarizes the descriptive statistics of the complete data sets.

This study considered one-step-ahead forecasting. Financial information is realized every month; therefore, we can adaptively adjust the model for next predictions. One-step-ahead forecasting can prevent problems associated with cumulative errors from the previous period that affect out-of-sample forecasting. The proposed model was trained in a batch manner; namely, 160 data points before the day of prediction were treated as the training data set, and the window of training data set slides with the current prediction, namely, moving the window one-step forward after every forecast and repeating the procedure. The daily returns in the final 60 days of the data series were used as the test data set for evaluating the performance of all prediction models. To measure experimental

Table 1. Data collection

Target Variable		Explanatory Variable	
Code	Variables	Code	Variables
GBP	Great British Pound	Y	Exchange rate (end of month)
JPY	Japanese Yen	CPI	Consumers Price Index
NTD	New Taiwan Dollar	IPI	Industry Product Index
DEM	Deutsche Mark	M	Money Supply M2
FRF	French Franc	R	Interbank Call Loan Rate
ITL	Italian Lira	L	Leading Indicator
		S	Stock Price Index
		$Trade$	Trade balance
		IFR	Inflator Rate.
		M_{US}	M2 (US)
		R_{US}	Interbank Call Loan Rate (US)

Table 2. Descriptive statistics for raw data

	GBP	JPY	NTD	DEM	FRF	ITL
Observations	221	221	221	221	221	221
Mean	0.0008	−0.0014	0.0008	−0.0004	−0.0004	0.0009
Median	0	−0.0007	0.0001	−0.0012	−0.0009	−0.0012
Maximum	0.1277	0.1139	0.073	0.0825	0.0827	0.1147
Minimum	−0.0846	−0.1547	−0.0458	−0.0661	−0.0655	−0.0655
Std. Dev.	0.0283	0.0321	0.0134	0.0258	0.0253	0.0266
Skewness	0.9688	−0.4526	0.6374	0.0749	0.1072	0.4455
Kurtosis	6.1298	5.5544	7.2991	3.1712	3.1449	4.0738

performance, the Root Mean of Squared Error (RMSE), the Mean Absolute Error (MAE) and the Directional Symmetry (DS) were employed. The definition of each indicator is as follows:

$$RMSE = \left[\frac{1}{N} \sum_{t=1}^{N} (Ypred_t - Y_t)^2 \right]^{\frac{1}{2}} \tag{18}$$

$$MAE = \frac{1}{N} \sum_{t=1}^{N} |Ypred_t - Y_t| \tag{19}$$

$$DS = \frac{1}{N} \sum_{t=1}^{N} ds_t \times 100\%$$

$$ds_t = \begin{cases} 1, & (Ypred_t - Y_{t-1}) \times (Y_t - Y_{t-1}) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

where $Ypred_t$ represents the forecasting values and Y_t represents the true values; in Formula (20), ds_t indicates if the prediction on price movement (up or down) is correct.

4.2. **Four pure forecasting models.** This study initially employed four pure models for forecasting, including GARCH (generalized autoregressive conditional heteroscedasticity), RBF-NN (radial basis function neural network), SVR (support vector regression) and RVR (Relevance Vector Regression). Figure 2 displayed the flow diagrams of the four pure models.

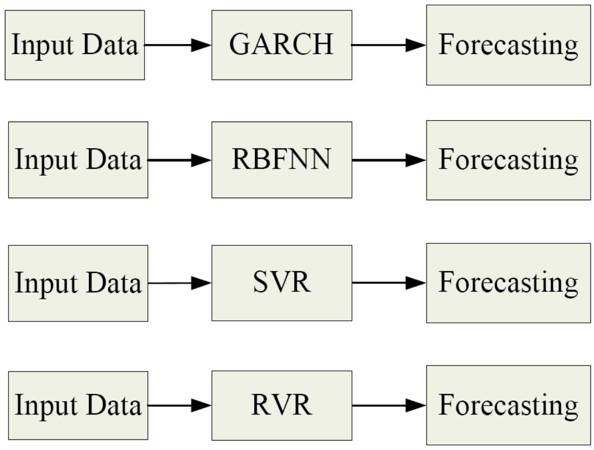

FIGURE 2. Flow diagrams of four pure models

4.3. **Performance evaluations on four pure models.** Tables 3 and 4 show the performance of four pure models. The conditions imposed to develop the results are as follows: (1) the input data for these models were the target exchange rate and macroeconomic factors. (2) For the GARCH model, a general GJR-GARCH $(1, 1)$ was set for the experiment. (3) The SVR model employed Gaussian kernel function for forecasting. The optimal parameters for the SVR model were searched by cross validation. (4) The RBF-NN used five neurons for the centers, and initialized via a k-means algorithm. The maximum iteration was 500 on training, the stopping criterion was an error smaller than 10^{-5}. (5) The RVR also used the Gaussian kernel. The initial parameters of RVR were $\sigma = 0.1$, $\alpha_i = 1$, $w_i = 0$, and the maximum iterations were 1000. Figure 2 illustrates the first stage experiment.

Table 3 shows the performance of GARCH under RMSE, MAE and DS. Table 4 shows the performance of RBF-NN, SVR and RVR. As listed in Tables 3 and 4, in pure models, RVR outperforms SVR and RBF-NN in each three criteria for all the return series, and SVR outperforms RBF-NN. The performance of RVR and GARCH is similar, but the RVR model is more effective and robust in DS predictions. These results confirm prior research (Tipping [37], Kim [22], Huang et al. [15], Ince and Trafalis [19], Hong et al. [16]).

TABLE 3. Forecasting results of GJR-GARCH $(1, 1)$ model

Exchange rate	RMSE	MAE	DS
GBP	0.0362	0.0245	68.33
JPY	0.0287	0.0225	75.00
NTD	0.0113	0.0080	86.67
DEM	0.0211	0.0155	76.67
FRF	0.0204	0.0159	75.00
ITL	0.0196	0.0147	75.00

4.4. **Performance comparison on wavelet-based models and the proposed hybrid model.** To reconstruct a phase space from observed data, it is necessary to confirm the chaotic nature of observed data and determine the embedding dimension m and delay time τ. The False Nearest Neighbors method (FNN, Kennel et al. [21]) and the Average Mutual Information method (AMI, Abarbanel [1]) were adapted to determine the embedding dimension (m) and delay time (τ) respectively. Table 5 presents the results of

TABLE 4. Forecasting results of RBF-NN, SVR and RVR

Exchange rate	RBF-NN			SVR			RVR		
	RMSE	MAE	DS	RMSE	MAE	DS	RMSE	MAE	DS
GBP	0.0256	0.0178	46.67	0.0235	0.0169	80.00	0.0232	0.0171	80.00
JPY	0.0292	0.0228	50.00	0.0286	0.0213	80.00	0.0277	0.0217	71.67
NTD	0.0129	0.0101	40.00	0.0136	0.0108	71.67	0.0132	0.0102	73.33
DEM	0.0325	0.0236	48.33	0.0263	0.0209	73.33	0.0259	0.0208	73.33
FRF	0.0702	0.0512	41.67	0.0259	0.0205	73.33	0.0262	0.0202	76.67
ITL	0.0264	0.0215	45.00	0.0264	0.0208	66.67	0.0263	0.021	70.00

optimal m and τ where the value denoted in boldface is the optimal dimension numbers. Consistent with the findings of Schwartz and Yousefis [35], the exchange rate returns in the current study belong to low-dimension chaos except for GBP, which is reconstructed by a five-dimensional space.

TABLE 5. Delay time and embedding dimension

m \ τ	**GBP/USD ($\tau = 3$)**	**DEM/USD ($\tau = 3$)**	**FRF/USD ($\tau = 1$)**	**NTD/USD ($\tau = 2$)**	**JPY/USD ($\tau = 2$)**	**ITL/USD ($\tau = 3$)**
1	73.2558	80.2326	77.9070	72.6744	77.9070	65.6977
2	23.8372	**9.3023**	**7.5581**	**12.2093**	**16.8605**	**8.7209**
3	3.4884	0.0000	0.0000	0.0000	0.0000	0.0000
4	1.1628	0.0000	0.0000	0.0000	0.0000	0.0000
5	**0.5814**	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: the value denoted in boldface in the table is the optimal dimension numbers.

After phase space reconstruction, this study employed the Daubechies wavelet with length 8 to decompose the phase space signal. Decomposed signals range from one low-frequency component (trend) and five high-frequency components (fluctuation); namely, the resolution level is five. Time-scale information was extracted from phase space signals through this step.

The procedure of the new model is shown in Figure 1. The conditions imposed to develop the results of Table 6 are as follows: (1) the optimal m and τ are displayed in Table 5. (2) The wavelet analysis used the Daubechies wavelet filter with length 8 to decompose the phase space signals. The resolution level was set to five. (3) A local RVR was used to forecast each frequency component. The final prediction was given by the combination of all frequency components. (4) The parameters used for RVR were the same as in Section 4.3.

For comparing the effectiveness of phase space representation, this study also performed wavelet-based SVR and RVR models for predictions. Figure 3 displays their flow diagrams. The conditions imposed to develop the experiment results of wavelet-based SVR and RVR are as follows: (1) the wavelet analysis also used the Daubechies wavelet filter with length 8 to decompose the input data. The resolution level was also set to five. 2) Different from our hybrid model, all frequency components were fed into a single (or global) SVR or

RVR for final forecasting. (3) The parameters used for RVR were the same as in Section 4.3. The SVR parameters were also searched by cross validation.

Table 6 exhibits the forecasting results of the wavelet-based models and our hybrid model (phase space reconstruction + wavelet decomposition + RVR), evaluated by each performance criteria. Comparing the two wavelet-based models with the new model, the new model obviously outperforms the wavelet-based models for each three criteria. The new model successfully obtained important state information of a chaotic dynamic system through reconstructing the phase space. Building prediction models in phase space is more effective and efficient.

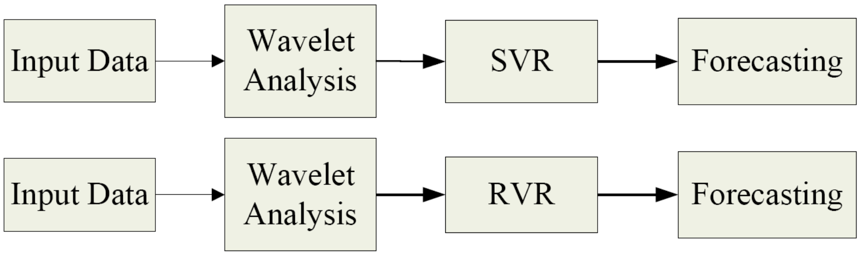

FIGURE 3. Flow diagrams of wavelet-based SVR and RVR models

TABLE 6. Forecasting results of wavelet-based models and our hybrid model

Exchange rate	Wavelet-SVR			Wavelet-RVR			Our hybrid model		
	RMSE	MAE	DS	RMSE	MAE	DS	RMSE	MAE	DS
GBP	0.0152	0.0156	88.33	0.0098	0.0113	88.33	0.0083	0.0065	90.00
JPY	0.0240	0.0211	78.33	0.0120	0.0113	91.67	0.0082	0.0061	96.67
NTD	0.0086	0.0084	88.33	0.0067	0.007	86.67	0.0045	0.0037	85.00
DEM	0.0164	0.0143	83.33	0.0106	0.0107	86.67	0.0070	0.0054	85.00
FRF	0.0151	0.0129	86.67	0.0138	0.0106	81.67	0.0072	0.0057	85.00
ITL	0.0157	0.0149	83.33	0.0126	0.0104	85.00	0.0072	0.0054	91.67

4.5. **Wilcoxon testing to confirm superiority of the new model.** This section summarizes performance under the three criteria and performs a formal statistical comparison of each model. Table 7 summarizes the results on RMSE; Table 8 summarizes the results on MAE; Table 9 summarizes the results on DS. On average, the pure SVR is poorer than the GJR-GARCH. This is because of the weakness of the pure SVR model in capturing volatility clustering, resulting in its underperformance compared to the GARCH type models. The RVR outperforms SVR and RBF-NN. RVR and SVR all have robust DS performance. The performance of RVR and GARCH is similar, but the RVR model is more effective and robust in DS predictions.

To confirm that our forecasting strategy is significantly better in statistics, this study conducted a Wilcoxon signed rank test (Wilcoxon [40]), a nonparametric alternative (for sample median) to the two sample t-test, based solely on the order in which the observations from the two samples fall. Table 10 shows the results on RMSE. Panel A of Table 10 reports the Wilcoxon tests of the GJR-GARCH model. Panel B of Table 10 displays the Wilcoxon comparison of our hybrid model with other models. Panel A of Table 10 shows that the wavelet-based models and our hybrid model significantly outperform (at 5%) the GJR-GARCH model. Panel B demonstrates that our hybrid model significantly outperforms (at 5%) all other models. This confirms that our hybrid model has the best performance.

Table 7. Comparison of RMSE for each model

Exchange rate	RBF-NN	SVR	RVR	Wavelet-SVR	Wavelet-RVR	GJR-GARCH	**Our hybrid model**
GBP	0.0256	0.0235	0.0232	0.0152	0.0098	0.0362	**0.0083**
JPY	0.0292	0.0286	0.0277	0.0240	0.0120	0.0287	**0.0082**
NTD	0.0129	0.0136	0.0132	0.0086	0.0067	0.0113	**0.0045**
DEM	0.0325	0.0263	0.0259	0.0164	0.0106	0.0211	**0.0070**
FRF	0.0702	0.0259	0.0262	0.0151	0.0138	0.0204	**0.0072**
ITL	0.0264	0.0264	0.0263	0.0157	0.0126	0.0196	**0.0072**
Average	0.03280	0.02405	0.02375	0.01583	0.01092	0.02288	**0.00707**

Table 8. Comparison of MAE for each model

Exchange rate	RBF-NN	SVR	RVR	Wavelet-SVR	Wavelet-RVR	GJR-GARCH	**Our hybrid model**
GBP	0.0178	0.0169	0.0171	0.0156	0.0113	0.0245	**0.0065**
JPY	0.0228	0.0213	0.0217	0.0211	0.0113	0.0225	**0.0061**
NTD	0.0101	0.0108	0.0102	0.0084	0.0070	0.0080	**0.0037**
DEM	0.0236	0.0209	0.0208	0.0143	0.0107	0.0155	**0.0054**
FRF	0.0512	0.0205	0.0202	0.0129	0.0106	0.0159	**0.0057**
ITL	0.0215	0.0208	0.0210	0.0149	0.0104	0.0147	**0.0054**
Average	0.02450	0.01853	0.0185	0.01453	0.01022	0.01685	**0.00547**

Table 9. Comparison of DS for each model

Exchange rate	RBF-NN	SVR	RVR	Wavelet-SVR	Wavelet-RVR	GJR-GARCH	**Our hybrid model**
GBP	46.67	80.00	80.00	88.33	88.33	68.33	**90.00**
JPY	50.00	80.00	71.67	78.33	91.67	75.00	**96.67**
NTD	40.00	71.67	73.33	88.33	86.67	86.67	**85.00**
DEM	48.33	73.33	73.33	83.33	86.67	76.67	**85.00**
FRF	41.67	73.33	76.67	86.67	81.67	75.00	**85.00**
ITL	45.00	66.67	70.00	83.33	85.00	75.00	**91.67**
Average	45.27833	74.167	74.167	84.721	86.668	76.111	**88.890**

Note: each forecasting value in the above table is evaluated on a percentage basis.

Table 10. Comparison of each models with Wilcoxon signed rank test

Exchange rates	RBF-NN	SVR	RVR	Wavelet-based SVR	Wavelet-based RVR	GJR-GARCH	Our hybrid model
Panel A: each value obtained from the comparison of GJR-GARCH and remainders with Wilcoxon signed rank test							
Stats	4	7	7	0**	0**	—	0**
p-value	0.2188	0.5625	0.5625	0.0313	0.0313	—	0.0313
Panel B: each value obtained from the comparison of our hybrid model and remainders with Wilcoxon signed rank test							
Stats	0**	0**	0**	0**	0**	0**	—
p-value	0.0313	0.0313	0.0313	0.0313	0.0313	0.0313	—

5. **Conclusions.** This study implements a novel model for exchange rate forecasting. The new model couples two useful techniques, phase space reconstruction and wavelet feature extraction, with Relevance Vector Regressions (RVRs) for nonparametric exchange

rate modeling and forecasting. By extracting nonlinear time series characteristics using these two techniques, the proposed model captures the movement of exchange rate returns more effectively, and consequently outperforms other forecasting models such as RBF-NN, SVRs, RVRs and GJR-GARCH.

Using AMI and FNN, this study obtained the optimal delay time and embedding dimension for phase space reconstruction. Based on this technique, this study transforms input space to phase space, which better captures the essential dynamics of a chaotic system. This study further applies wavelet decomposition to map the phase space signal into several time-frequency domains to forecast individually by different local models. Each time-frequency component is finally fed into RVRs for prediction. Experimental results on six exchange rate returns measured by three performance criteria prove that these two feature extraction techniques powerfully reduce forecasting errors. Compared with general wavelet-based and traditional time series models, our hybrid model performs best and obtains a statistically significant performance improvement.

5.1. **Limitations of the study and suggestions for future research.** This investigation has certain limitations that need to be considered. Some of these limitations may represent fruitful avenues for future studies. First, this study employs a traditional wavelet decomposition, which uses a fixed basis to perform wavelet transformation. To adapt to any types of data, future studies may consider adaptive wavelet transformations via lifting or other schemes. Second, this study constructs a local predictor on each wavelet component, which is computational intensive and more suitable to operate on a parallel or distributed computing environment. In contrast, future research may consider a global model in the wavelet domain. Third, semi-supervised learning is a new technique to improve forecasting performance. Finally, trading is also an important issue in the financial application of soft computing. How to effectively utilize the forecasting power of this study for trading requires further investigation.

Acknowledgment. The authors gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data*, Springer, New York, USA, 1996.

[2] M. Abdollahzade, M. J. Mahjoob, R. Zarringhalam and A. Miranian, Locally linear neuro-fuzzy (LLNF) eelectricity price forecasting in deregulated power markets, *International Journal of Innovative Computing, Information and Control*, vol.6, no.9, pp.4203-4218, 2010.

[3] N. Begum, M. A. Fattah and F. Ren, Automatic text summarization using support vector machine, *International Journal of Innovative Computing, Information and Control*, vol.5, no.7, pp.1987-1996, 2009.

[4] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, vol.31, pp.307-327, 1986.

[5] J. F. Chang, L. Y. Wei and C. H. Cheng, ANFIS-based adaptivce expectation model for forecasting stock index, *International Journal of Innovative Computing, Information and Control*, vol.5, no.7, pp.1949-1958, 2009.

[6] R. C. Chen and S. P. Chen, Intrusion detection using a hybrid support vector machine based on entropy and TF-IDF, *International Journal of Innovative Computing, Information and Control*, vol.4, no.2, pp.413-424, 2008.

[7] S. Y. Chen and K. Jeong, Recurrent support vector regression for a nonlinear ARMA model with applications to forecasting financial returns, *Sonderforschungsbereich 649 Discussion Papers*, 2008.

[8] G. Chen, T. J. Wang, L. Y. Gong and P. Herrera, Multi-class support vector machine active learning for music annotation, *International Journal of Innovative Computing, Information and Control*, vol.6, no.3(A), pp.921-930, 2010.

[9] G. Coban and A. H. Buyuklu, Deterministic flow in phase space of exchange rates: Evidence of chaos in filtered series of Turkish Lira-Dollar daily growth rates, *Chaos, Solitons and Fractals*, vol.42, pp.1062-1067, 2009.
[10] M. Z. Frank and T. Stengos, Some evidence concerning macroeconomic chaos, *Journal of Monetary Economics*, vol.22, pp.423-438, 1988.
[11] V. Fernandez, Wavelet- and SVM-based forecasts: An analysis of the U.S. metal and materials manufacturing industry, *Resources Policy*, vol.32, pp.80-89, 2007.
[12] M. Gallegati, Wavelet analysis of stock returns and aggregate economic activity, *Computational Statistics & Data Analysis*, vol.52, pp.3061-3074, 2008.
[13] L. Glosten, R. Jagannathan and D. Runkle, Relationship between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, vol.48, pp.1779-1801, 1993.
[14] M. Han, W. Guo and M. Fan, Multivariate chaotic time series prediction based on radial basis function neural network, *International Symposium on Neural Networks*, pp.741-746, 2006.
[15] W. Huang, Y. Nakamori and S. Y. Wang, Forecasting stock market movement direction with support vector machine, *Computers & Operations Research*, vol.32, pp.2513-2522, 2005.
[16] W. C. Hong, Y. F. Chen, P. W. Chen and Y. H. Yeh, Continuous ant colony optimization algorithms in a support vector regression based financial forecasting model, *Computer Society*, vol.1, pp.548-552, 2007.
[17] F. In and S. Kim, Multiscale hedge ratio between the Australian stock and futures markets: Evidence from wavelet analysis, *Journal of Multinational Financial Management*, vol.16, pp.411-423, 2006.
[18] M. Iseri, H. Caglar and N. Caglar, A model proposal for the chaotic structure of Istanbul stock exchange, *Chaos, Solitons and Fractals*, vol.36, no.5, pp.1392-1398, 2008.
[19] H. Ince and T. B. Trafalis, A hybrid model for exchange rate prediction, *Decision Support Systems*, vol.42, pp.1054-1062, 2006.
[20] J. Karuppiah and C. A. Los, Wavelet multiresolution analysis of high-frequency Asian FX rates, *International Review of Financial Analysis*, vol.14, pp.211-246, 2005.
[21] M. B. Kennel, R. Brown and H. D. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction, *Physical Review Letter*, vol.45, no.6, pp.3403-3411, 1992.
[22] K. J. Kim, Financial time series forecasting using support vector machines, *Neurocomputing*, vol.55, pp.307-319, 2003.
[23] S. Kim and F. In, The relationship between stock returns and inflation: New evidence from wavelet analysis, *Journal of Empirical Finance*, vol.12, pp.435-444, 2005.
[24] S. Kim and F. In, On the relationship between changes in stock prices and bond yields in the G7 countries: Wavelet analysis, *Journal of International Financial Markets, Institutions and Money*, vol.17, pp.167-179, 2007.
[25] M. Kurban and U. B. Filik, Next day load forecasting using artificial neural network models with autoregression and weighted frequency bin blocks, *International Journal of Innovative Computing, Information and Control*, vol.5, no.4, pp.889-898, 2009.
[26] K. W. Lau and Q. H. Wu, Local prediction of non-linear time series using support vector regression, *Pattern Recognition*, vol.41, pp.1539-1547, 2008.
[27] D. F. Li, W. C. Hu, W. Xiong and J. B. Yang, Fuzzy relevance vector machine for learning from unbalanced data and noise, *Pattern Recognition Letters*, vol.29, no.9, pp.1175-1181, 2008.
[28] F. Lisi and R. Schiavo, A comparison between neural networks and chaotic models for time series prediction, *Computational Statistics and Data Analysis*, vol.30, pp.87-102, 1999.
[29] L. L. Ma and X. S. Xu, RBF network-based chaotic time series prediction and it's application in foreign exchange market, *Proc. of 2007 International Conference on Intelligent Systems and Knowledge Engineering*, 2007.
[30] S. Mitra, A wavelet filtering based analysis of macroeconomic indicators: The Indian evidence, *Applied Mathematics and Computation*, vol.17, pp.1055-1079, 2006.
[31] N. H. Packard, J. P. Crutchfield, J. D. Farmer and R. S. Shaw, Geometry from a time series, *Physical Review Letter*, vol.45, pp.712-716, 1980.
[32] F. Perez-Cruz, J. Afonso-Rodriguez and J. Giner-Rubio, Estimating GARCH models using support vector machines, *Quantitative Finance*, vol.3, pp.1-10, 2003.
[33] A. Rua and L. C. Nunes, International comovement of stock market returns: A wavelet analysis, *Journal of Empirical Finance*, vol.16, pp.632-639, 2009.
[34] J. A. Scheindman and B. Lebaron, Nonlinear dynamics and stock returns, *The Journal of Business*, vol.62, pp.311-337, 1989.

[35] B. Schwartz and S. Yousefi, On complex behavior and exchange rate dynamics, *Chaos, Solitons and Fractals*, vol.18, no.3, pp.503-523, 2003.

[36] F. Takens, Detecting strange attractors in turbulence, *Lecture Notes in Mathematics*, vol.898, pp.366-381, 1981.

[37] M. E. Tipping, Sparse Bayesian learning and the relevance vector machine, *Journal of Machine Learning Research*, vol.1, pp.211-244, 2001.

[38] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, USA, 1995.

[39] H. C. Wang, G. R. Chen and J. H. Lu, Complex dynamical behaviors of daily data series in stock exchange, *Physics Letters A*, vol.333, pp.246-255, 2004.

[40] F. Wilcoxon, Individual comparisons by ranking methods, *Biometrics*, vol.1, pp.80-83, 1945.

[41] H. Wu and X. Chang, Power load forecasting with least squares support vector machines and chaos theory, *Proc. of the 6th World Congress on Control and Automation*, pp.4369-4373, 2006.

[42] C. Y. Yeh, C. W. Huang and S. J. Lee, A multiple-kernel support vector regression approach for stock market price forecasting, *Expert Systems with Applications*, vol.38, no.3, pp.2177-2186, 2011.

[43] S. Yousefi, I. Weinreich and D. Reinarz, Wavelet-based prediction of oil prices, *Chaos, Solitons and Fractals*, vol.25, pp.265-275, 2005.

[44] B. L. Zhang, R. Coggins, M. A. Jabri, D. Dersch and B. Flower, Multiresolution forecasting for futures trading using wavelet decompositions, *IEEE Transactions on Neural Networks*, vol.12, no.4, pp.765-775, 2001.

[45] Z. Zhang, H. Ikeuchi, N. Saiki, T. Imamura, T. Miyake, H. Toda and S. Horihata, Fast wavelet instantaneous correction and its application to abnormal signal dectection, *International Journal of Innovative Computing, Information and Control*, vol.4, no.10, pp.2697-2710, 2008.