

## Research Article

# Exchange Rate Forecasting Using Entropy Optimized Multivariate Wavelet Denoising Model

Kaijian He,<sup>1,2</sup> Lijun Wang,<sup>1</sup> Yingchao Zou,<sup>1</sup> and Kin Keung Lai<sup>2,3</sup>

<sup>1</sup> School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029, China

<sup>2</sup> Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Kowloon, Hong Kong

<sup>3</sup> School of Economics and Management, North China Electric Power University, Beijing 102206, China

Correspondence should be addressed to Kaijian He; [kaijian.he@my.cityu.edu.hk](mailto:kaijian.he@my.cityu.edu.hk)

Received 29 December 2013; Accepted 19 March 2014; Published 17 April 2014

Academic Editor: Fenghua Wen

Copyright © 2014 Kaijian He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Exchange rate is one of the key variables in the international economics and international trade. Its movement constitutes one of the most important dynamic systems, characterized by nonlinear behaviors. It becomes more volatile and sensitive to increasingly diversified influencing factors with higher level of deregulation and global integration worldwide. Facing the increasingly diversified and more integrated market environment, the forecasting model in the exchange markets needs to address the individual and interdependent heterogeneity. In this paper, we propose the heterogeneous market hypothesis- (HMH-) based exchange rate modeling methodology to model the micromarket structure. Then we further propose the entropy optimized wavelet-based forecasting algorithm under the proposed methodology to forecast the exchange rate movement. The multivariate wavelet denoising algorithm is used to separate and extract the underlying data components with distinct features, which are modeled with multivariate time series models of different specifications and parameters. The maximum entropy is introduced to select the best basis and model parameters to construct the most effective forecasting algorithm. Empirical studies in both Chinese and European markets have been conducted to confirm the significant performance improvement when the proposed model is tested against the benchmark models.

## 1. Introduction

In the post-Bretton Woods era, the worldwide exchange markets have shifted towards the more floating and volatile era, which are characterized by high level of fluctuations and risk exposures. Given its role as one of the most important economic factors for the national economy in the increasingly open and globalized economic system, the accurate and reliable forecasting of the exchange rate movement has profound impacts throughout different levels of the economy, including government, enterprises, and academics [1].

Theoretically numerous empirical studies have been conducted to investigate the interrelationship and comovement between the exchange markets and other markets, including crude oil market, stock market, and bond market. For example, Salisu and Mobolaji [2] found the statistical evidence of bidirectional relationship between oil price and US-Nigeria

exchange rate [2]. Chkili and Nguyen [3] use a regime-switching model to identify the dynamic linkages between the exchange rates and stock market returns during both calm and turbulent periods [3]. Hacker et al. [4] have found new evidence of negative linkage between the exchange rate and interest rate differentials in the wallet time scale domain [4]. Meanwhile in the literature there is much less attention paid to the exploration of linkage among different exchange markets, which represent the essential theoretical challenge. For example, Kim et al. [5] find that the conditional correlation between Japanese Yen and other Asian economies is decoupled and insignificant due to liquidity deterioration and elevated risk aversions in the international capital market [5]. But recent empirical studies suggest that the comovement across markets, with transmission as one particular case, turns out to be more complicated than what is assumed in the traditional linear framework. This may stem from the loss of information using the low frequency data and the

inferior methodology to analyze more accurately the cross correlations.

Methodologically traditional linear structural models have been effective in the forecasting of the exchange rate movement over the medium to long time horizon, with acceptable approximation accuracy and computational efficiency, where the aggregated price behavior is comparably stable and stationary. These include models in both asset and monetary views as well as Keynesian view, regression models, cointegration and vector autoregressive models, multivariate stochastic models, and so forth [6]. When it comes to the shorter time horizon, these traditional models also have largely failed to demonstrate the competent forecasting performance in the empirical studies [6]. Part of the reasons can be attributed to the nonlinear data characteristics revealed in the recent empirical studies [7, 8]. Different distribution free artificial intelligence techniques such as neural network and support vector regressions, together with the innovative optimization methods, have shown superior performance under different circumstances [7, 9–11]. But they are mainly black box in nature and offer little insights into the underlying patterns as well as supporting theories behind.

Recent empirical studies on the fractal and multiscale data characteristics indicate the emergence of multiscale modeling as the important alternative [7]. Wavelet analysis, as one popular multiscale modeling technique, has been introduced to model not only horizontal dependency in the time domain such as volatility clustering (conditional heteroscedasticity) and long memory (slow decaying autocorrelation) but also vertical dependency across time domain simultaneously [12, 13]. For example, Tiwari et al. [14] identify the comovement of oil price and Indian Rupee at higher time scales, but not lower ones, using wavelet analysis [14]. Reboredo and Rivera-Castro [15] use the wavelet analysis to disentangle the oil price-exchange rate relationship in the time scale domain [15]. Orlov [16] tests for the time varying exchange rate comovements at different time scales [16]. Recently the emergence of this approach leads to the emergence of the heterogeneous market hypothesis (HMH) as the theoretical foundation to replace efficient market hypothesis (EMH) to model the heterogeneous data characteristics. Methodologically this approach helps explain the heterogeneity of the underlying data comovement and transmission mechanism, behind the exhibited nonlinear data characteristics. However, in the exchange rate forecasting literature, we have only witnessed limited attempts in modeling the correlations and comovements among exchange markets when constructing the forecasting algorithm.

In the meantime, the entropy theory has been used to analyze the information content of the wavelet decomposed multiscale data structure. Wavelet entropy, relative wavelet entropy, and many other variants have been proposed in the literature to calculate the entropy of the energy distribution in the typical wavelet decomposition, as well as the cost function for the best basis algorithm to choose the optimal basis for wavelet packet transform [17, 18]. Xu et al. [19] use the modified wavelet entropy measure to differentiate between the normal and hypertension states [19]. Samui and Samantaray [20] incorporate the wavelet entropy measure in constructing

the measuring index for islanding detection in distributed generation [20]. Wang et al. [21] use best basis-based wavelet packet entropy to extract feature in the decomposed structure for the follow-up classification algorithm, which performs well in EEG analysis for patient classification [21]. In the forecasting field, the entropy maximization has recently been proposed to select the best forecasters, but with much less attention attracted in the literature. For example, Bessa et al. [22] adopt the maximum entropy criteria in neural network training and find it to provide more superior performance than traditional mean square error (MSE) criteria in wind power prediction [22].

In this paper we propose an innovative entropy optimized multivariate wavelet denoising model. Empirical studies are conducted in the closely related Chinese and European exchange markets to evaluate the additional value offered by the incorporation of nonlinear multiscale cross-markets correlations in the proposed algorithm. Our contributions are threefold. Firstly we provide the empirical evidence of multiscale heterogeneous data characteristics distinguishable by sizes. Secondly we incorporate this stylized fact in the construction of the innovative wavelet denoising-based forecasting algorithms. Thirdly we propose the maximum entropy as a measure for in-sample performance to select the best basis and decomposition level. To the best of our knowledge, work in this paper is unique and amongst the first in introducing the maximum entropy in forecasting the exchange rate movement in the multiscale domain, to select the best basis and parameters.

The rest of the paper proceeds as follows. Section 2 briefly reviewed the two relevant theories, that is, multivariate wavelet denoising theory and the entropy theory, underlying the proposed model. Section 3 proposes the multivariate wavelet analysis to analyze the time varying correlations. We further construct the multivariate wavelet denoising-based exchange rate forecasting algorithm. In Section 4 we conduct experiments to empirically test and confirm the performance superiority of the proposed algorithm against the benchmark models. Detailed analysis of experiment results is illustrated as well. Section 5 concludes with summarizing remarks.

## 2. Multivariate Wavelet-Based Denoising Theory and Entropy Theory

The ultimate goal of denoising is to set the right boundary and remove the noises while preserving major data features. In recent years wavelet denoising algorithm dominates more traditional methods such as moving average filter, exponential smoothing filter, linear Fourier smoothing, and simple nonlinear noise reduction, as it does not assume homogeneous error structure. For example, Kwon et al. [23] point out the problems with existing denoising techniques as assuming homogeneous error structure and they proposed wavelet denoising method incorporating a variance change point detection thresholding method to deal with it in protein mass spectroscopy applications [23]. Boto-Giralda et al. [24] use the stationary wavelet-based denoising methods to improve the performance of traffic volume prediction models in

intelligent transportation system [24]. Gao et al. [25] propose an adaptive denoising algorithm and contend it to be superior to wavelet-based approaches when applied to analysis of electroencephalogram (EEG) signals contaminated with noises [25]. Lotric and Dobnikar [26] and Lotric [27] integrate the neural network with the wavelet denoising method to optimize the denoising parameters dynamically and find the performance improvement in prediction accuracy [26, 27].

Generally the multivariate wavelet denoising algorithm involves the following procedures.

(1) The original data series are projected into the general higher-dimensional space  $L_2(\mathbb{R}^d)$  into different subspaces characterized by scales using multivariate wavelet transform. These subspaces constitute a doubly infinite nested sequence of subspaces pairs of both denoised and noise data components  $(V_j^{\text{DN}}, V_j^{\text{N}})$ , respectively, of  $L_2(\mathbb{R}^d)$  as follows [28–30]:

$$\begin{aligned} \cdots &\subset (V_2^{\text{DN}}, V_2^{\text{N}}) \subset (V_1^{\text{DN}}, V_1^{\text{N}}) \subset (V_0^{\text{DN}}, V_0^{\text{N}}) \subset (V_{-1}^{\text{DN}}, V_{-1}^{\text{N}}) \\ &\subset (V_{-2}^{\text{DN}}, V_{-2}^{\text{N}}) \subset \cdots \subset L_2(\mathbb{R}^d) \cap_{j \in \mathbb{Z}} (V_j^{\text{DN}}, V_j^{\text{N}}) = \{0\}, \\ &\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L_2(\mathbb{R}^d) \end{aligned} \quad (1)$$

in which

$$(V_j^{\text{DN}}, V_j^{\text{N}}) = \bigotimes_{i=1}^d (V_{j,i}^{\text{DN}}, V_{j,i}^{\text{N}}) \subset L_2(\mathbb{R}^d), \quad (2)$$

where  $\bigotimes$  is the tensor product operator.

Then for each  $(V_{j-1}^{\text{DN}}, V_{j-1}^{\text{N}})$  in  $(V_j^{\text{DN}}, V_j^{\text{N}})$ , the orthogonal complement space  $(W_j^{\text{DN}}, W_j^{\text{N}})$  is defined as

$$(V_{j-1}^{\text{DN}}, V_{j-1}^{\text{N}}) = (V_j^{\text{DN}}, V_j^{\text{N}}) \bigoplus (W_j^{\text{DN}}, W_j^{\text{N}}). \quad (3)$$

For all  $f, g \in \mathbb{L}^2(\mathbb{R}^2)$ , the decomposition and reconstruction in the two-dimensional case is defined as in (4), respectively:

$$\begin{aligned} \int_0^\infty \frac{da}{a^{n+1}} \int_{-\infty}^\infty db (T^{\text{wav}} f)(a, b) \overline{(T^{\text{wav}} g)(a, b)} &= C_\psi \langle f, g \rangle, \\ f &= C_\psi^{-1} \int_0^\infty \frac{da}{a^{n+1}} \int_{\mathbb{R}^n} db (T^{\text{wav}} f)(a, b) \psi^{a,b}, \end{aligned} \quad (4)$$

where  $(T^{\text{wav}} f)(a, b) = \langle f, \psi^{a,b} \rangle$ ,  $\psi^{a,b}(x) = a^{-n/2} \psi((x-b)/a)$ , and  $a \in \mathbb{R}_+$ ,  $a \neq 0$ ,  $b \in \mathbb{R}^n$ . In the 2-dimensional case, three orthonormal wavelet bases, including horizontal wavelet  $\Psi^{\text{horizontal}}(x, y)$ , vertical wavelet  $\Psi^{\text{vertical}}(x, y)$ , and diagonal wavelet  $\Psi^{\text{diagonal}}(x, y)$ , are needed to produce subspaces [31].

When the data are of finite support, the discrete wavelet transform would encounter the boundary distortion issue at the edge of the data under analysis. Padding, that is, adding extra data points to either left or last data, is one approach to facilitate the transform. Different padding techniques exist, including zero padding and symmetric padding [30].

(2) The separation between subspaces pairs  $(V_j^{\text{DN}}, V_j^{\text{N}})$  and  $(W_j^{\text{DN}}, W_j^{\text{N}})$  is achieved by applying the threshold chosen specifically at different scales for different directions to either suppress or shrink the wavelet coefficients. The denoised and noise part are separated with finer details revealing patterns at more microscales.

The dominant threshold selection rules include Universal, Minimaxi, and Steins unbiased risk estimate (SURE) [32]. Different threshold selection rules have different targets when setting the noise reduction target. For example, the universal threshold selection rule aims to reduce the noises at maximum level possible. The threshold is selected as  $\eta^U = \hat{\sigma} \sqrt{2 \log N}$ , where  $N$  is the number of wavelet coefficients and  $\hat{\sigma}$  is the estimate of the volatility level. When the sample size  $N$  is large and the data series are normally distributed, the universal threshold gives the upper bound value to the noise level in the data statistically. However, this method achieves the maximal level of smoothness at the cost of lower goodness-of-fit. The denoised data risk loses some important data features. The Minimaxi threshold selection rule adopts the function fitness criteria such as MSE. The denoised data represent the best fit approximation to the original data, retaining spikes and hicks. However, this is achieved at the cost of lower function smoothness.

The mainstream shrinkage rules are hard and soft threshold selection rules [32, 33]. The hard threshold selection rule is the high pass filter, which suppresses the wavelet coefficients below the chosen threshold values and leaves the rest coefficients intact as follows:

$$\delta_\eta^H(O_t) = \begin{cases} O_t & \text{if } |O_t| > \eta \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

where  $O_t$  refers to the wavelet coefficients and  $\eta$  is the set threshold value. The soft shrinkage rules focus on the signal smoothing. It suppresses the wavelet coefficients below the set threshold value and subtracts the threshold value from the remaining wavelet coefficients. Compared to hard threshold selection rules, the data processing following soft threshold selection rules is smoother but loses the abrupt changes in the original data. It filters the signal as follows:

$$\delta_\eta^S(O_t) = \text{sign}(O_t) (|O_t| - \eta)_+, \quad (6)$$

where

$$\text{sign}(O_t) = \begin{cases} +1 & \text{if } O_t > 0 \\ 0 & \text{if } O_t = 0 \\ -1 & \text{if } O_t < 0, \end{cases} \quad (7)$$

$$(x)_+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

(3) Processed wavelet coefficients are reconstructed into the unified data series using wavelet synthesis.

During the denoising process, there are unknown parameters that have significant impacts on the denoising performance. Other than the statistical approach to quantify

them, the information theoretic approach can be brought in to quantify them as well. The entropy is a widely used statistical measure of disorder and uncertainty, quantifying the data randomness [34]. It also corresponds to a measure of information content. For a stochastic time series system, the classical Shannon entropy is defined as follows [35]:

$$S = \int_0^\infty -f(x) \ln(f(x)) dx. \quad (8)$$

The value of entropy lies between 0 and 1. The higher the entropy is, the higher the level of disorder and uncertainty is.

### 3. An Entropy Optimized Multivariate Wavelet Denoising-Based Exchange Rate Forecasting Model (MWVAR)

Homogeneity and rationality are two basic assumptions imposed in the traditional EMH behind major multivariate exchange rate forecasting algorithms. To recognize the multiscale properties in the high frequency data, we propose the HMH instead. In HMH, the heterogeneous market microstructure is acknowledged explicitly, by assuming different investors strategy, scale, and time horizon, just to name a few [36–42].

Following HMH framework, the exchange market receives the joint influence from market agents with different defining characteristics including investment strategies, time horizons, and investment scales.

Based on the stylized facts, we make some simplifying assumptions: (1) investment strategies within each time horizon are homogeneous and (2) investment strategies across time scales are mutually independent. Then based on the aforementioned theoretical framework, we propose the entropy optimized multivariate wavelet denoising-based exchange rate forecasting algorithm. It involves the following steps.

(1) Suppose that the return of the exchange markets are the sum of common latent factors and the individual latent factors; the multivariate wavelet denoising algorithm is used to separate data from noise using particular wavelet families. By decomposing data into the multiscale domain, data and noises are separated based on their different characteristics across scales with noise smaller in scales. Thus more subtle distinction between data and noise can be set.

(2) The denoised and noise data are supposed to be following some particular stochastic process. The conditional mean matrix for the denoised data and noise data is modeled by employing the particular conditional time series models. In this paper, we adopt vector autoregressive (VAR) processes as follows:

$$\hat{y}_t = \delta_t + \sum_{i=1}^m \phi_i y_{t-i} + \sum_{j=1}^n \theta_j \varepsilon_{t-j} + \varepsilon_t, \quad (9)$$

where  $\hat{y}_t$  is the conditional mean at time  $t$ ,  $y_{t-i}$  is the lag  $m$  returns with parameter  $\phi_i$ , and  $\varepsilon_{t-j}$  is the lag  $n$  residuals in the previous period with parameter  $\theta_j$ .

VAR is chosen over more theoretically sound vector autoregressive moving average (VARMA) model in this

paper due to the following reasons: firstly there is lack of authoritative methodology to uniquely identify and estimate VARMA model, although some initial attempts have been made [43]. VAR is still by far the most well established and applied multivariate time series models in the literature. Secondly any invertible vector ARMA can be approximated by VAR with infinite order [44]. We use the information criteria (IC) to determine the optimal specification for the VAR model. Typical IC includes akaike information criteria (AIC) and Bayesian information criteria (BIC).

(3) Using the in-sample data, different criteria such as MSE and entropy can be used to determine the model specifications and parameters. The minimization of MSE corresponds to the minimization of error variance. The entropy maximization corresponds to the maximization of information content in the predictors and higher generalizability. Given the predicted random variable  $\hat{Y} \in R^n$ , generated with the unknown data generating process (DGP) with unknown parameters and the observation  $Y \in R^n$ , the Shannon entropy of predictor is defined as follows:

$$H(\hat{Y}) = E[-\log p(\hat{Y})] = - \int_{R^n} p(Y) \log Y(Y) dx, \quad (10)$$

where  $H(\hat{Y})$  refers to the Shannon entropy of the predictor  $\hat{Y}$ ,  $p(x)$  refers to the probability density function (PDF). The objective is to maximize the  $H(\hat{Y})$  of the measurable function  $Y$  by adjusting different parameters of forecasting algorithm that produces  $\hat{Y}$ .

(4) With the chosen model specifications and parameters, the forecast matrix is reconstructed from the individual forecasts, both denoised and noise parts. Thus the mean matrix can be aggregated from the individual mean matrix forecasts.

## 4. Empirical Studies

**4.1. Experiment Settings.** We choose the US Dollar against Chinese Renminbi (RMB) and US Dollar against European Euro (Euro) exchange rates to construct the data set for the empirical studies. The dataset in the empirical studies extends from 23 July 2007 to 30 August 2013. The starting date is chosen as the Chinese government changes its exchange rate policy from the fixed pegging to US Dollar to a basket of currency. This has significant impact on the exchange rate movement, shifting into a different regime and evolving based on different underlying mechanism with wider bands as well as higher level of fluctuations. The end date is set to include the latest data available when the research was conducted. The dataset is preprocessed for recording errors and to remove the records when the readings are interrupted due to holiday breaks in either market. This results in 1533 daily observations. Since there is no consensus on the division of dataset, either in the machine learning or econometric literature [45], when dividing the dataset, we follow the common criteria that reserves at least 60% data as the training set and retain sufficiently large size of the test set for the results to be statistically valid [46]. The dataset is divided into three subdataset, that is, the training set for the proposed wavelet



TABLE 1: Descriptive statistics and statistical tests.

Statistics	$P_{\text{Euro}}$	$P_{\text{RMB}}$	$r_{\text{Euro}}$	$r_{\text{RMB}}$
Mean	1.3706	0.1503	0	0.0001
Max	1.6010	0.1636	0.0461	0.0059
Min	1.1957	0.1315	-0.0303	-0.0087
Std	0.0877	0.0079	0.0071	0.0011
Skewness	0.5550	-0.2730	0.2650	-0.1982
Kurtosis	2.6664	2.4969	5.6767	9.4280
$p_{\text{JB}}$	0.0010	0.0010	0.0010	0.0010

denoising VAR model (36%), the model tuning set to select the best basis and relevant parameters (24%), and the test set for the out-of-sample test to evaluate the performance of different models (40%). We perform one day ahead forecast using rolling-window method.

Descriptive statistics for data characteristics are listed in Table 1.

Where  $P_{\text{Euro}}$  and  $P_{\text{RMB}}$  refer to the price of both euro and RMB.  $p_{\text{JB}}$  refers to the  $P$  value of the JB test statistics. Descriptive statistics in Table 1 show some interesting stylized facts. The market exhibits considerable fluctuations, as suggested by the significant volatility level. The distribution of the market price is fat-tail and leptokurtic, as suggested by significant skewness and kurtosis levels. There is also high level of market risk exposure due to extreme events in the market, as reflected in the significant kurtosis level. The market return also deviates from the normal distribution and exhibits nonlinear dynamics; this is further confirmed by the rejection of Jarque-Bera test of normality [47, 48]. As autocorrelation and partial autocorrelation function indicate the trend factors, the daily prices are log differenced at the first order to remove them as in  $y_t = \ln(p_t/(p_{t-1}))$ . We further calculate the descriptive statistics on the return data and find it to approximate the normal distribution, as indicated by the four moments. The kurtosis appears to deviate from the normal level, which indicates that the market exhibits significant abnormal return changes event. Besides, since the null hypothesis of JB test is rejected, this further indicates that the market return contains unknown nonlinear dynamics, not easily captured by traditional linear models.

**4.2. Empirical Analysis of Dynamic Behaviors.** The exchange markets are subject to increasingly frequent and abrupt external shocks such as the subprime crisis and European debt crisis, which are illustrated in the plot of returns of both Chinese Renminbi and European Euro as in Figures 1 and 2.

This is where  $r_{\text{rmb},i}$ ,  $i = 1, 2, 3$  refers to the RMB returns over period  $i$ .  $r_{\text{euro},i}$ ,  $i = 1, 2, 3$  refers to the Euro returns over period  $i$ . The three periods refer to July, 2007 to May, 2009; May, 2009 to December, 2011; and December, 2011 to August, 2013. It can be seen from Figures 1 and 2 and Table 2 that returns in both markets share some common features while demonstrating their unique characteristics. Both markets are subject to the influences of the subprime crisis from February, 2007 and spread to May, 2009 and European debt crisis from December, 2009 to December, 2011. Meanwhile individually

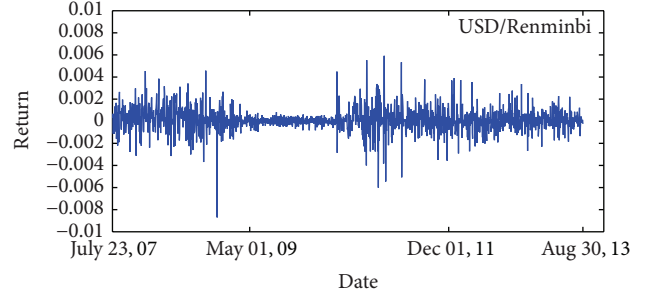


FIGURE 1: Returns of USD/Renminbi.

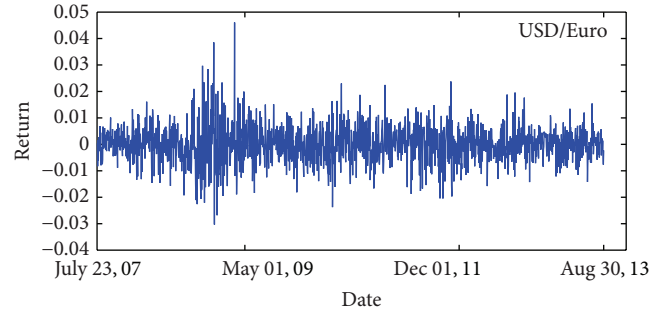


FIGURE 2: Returns of USD/Euro.

each market is subject to the influences of some unique external forces. As shown in Figure 1, Renminbi market has much lower level of fluctuations than the Euro markets, subject to tighter control of the central bank. As shown in Figure 2, the European debt crisis has much stronger impact on the Euro markets than Renminbi market. Generally the impact of subdebt crisis is larger than the European debt crisis for both markets. Meanwhile, as shown in both figures, both markets are subject to some common latent factors. Both markets demonstrate the obvious jump behaviours around 2008, when they are subject to the subprime crisis.

We further calculate the cross correlations between two exchange markets, using the roll windows of size 252 and plotting the results in Figure 3.

It can be seen from Figure 3 that the correlations are dynamically changing, over different periods at the significant level. Taking subprime crisis and European debt crisis as the marking events, we further calculated the descriptive statistics of correlations for three periods and listed the results in Table 3.

It can be seen from both Figure 3 and Table 3 that for periods 1 and 3, the correlation is positive and relatively stable. The correlations turn negative for period 2, accompanied by significant fluctuations. Within period two, there are frequent regime-switching behaviors. For example, the correlation from November, 2007 to August, 2009 is mainly affected by subprime crisis. The other result from European debt crisis is as follows. We see that from November, 2009 to December, 2010, the correlation stays at the positive level, subject to the positive measures taken by the European Union. Then, during December, 2010 and August, 2011, the

TABLE 2: Descriptive statistics and statistical tests.

Statistics	$r_{\text{rmb},1}$	$r_{\text{rmb},2}$	$r_{\text{rmb},3}$	$r_{\text{euro},1}$	$r_{\text{euro},2}$	$r_{\text{euro},3}$
Minimum	-0.0087	-0.0060	-0.0029	-0.0303	-0.0237	-0.0158
Maximum	0.0046	0.0059	0.0039	0.0461	0.0238	0.0196
Mean	0.0002	0.0001	0.0001	-0.0001	0	0
Standard deviation	0.0012	0.0011	0.00094435	0.0087	0.0070	0.0054
Kurtosis	9.2450	10.7865	5.1886	6.2843	3.4396	3.7204
Skewness	-0.7236	-0.0094	0.4462	0.5609	-0.0970	0.1224
$p_{\text{JB}}$	0.0010	0.0010	0.0010	0.0010	0.0434	0.0109

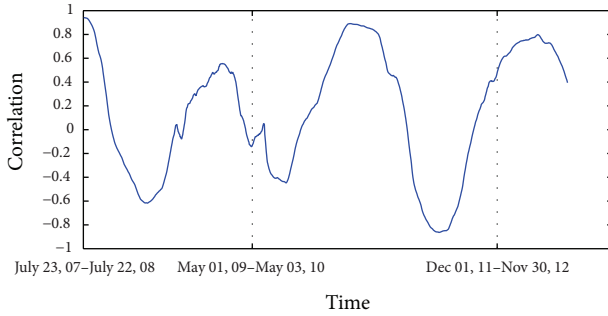


FIGURE 3: Dynamic correlations between USD/RMB and USD/Euro.

TABLE 3: Descriptive statistics of cross correlations.

Statistics	$\rho_1$	$\rho_2$	$\rho_3$
Minimum	-0.6171	-0.4471	0.3941
Maximum	0.9417	0.8911	0.7987
Mean	-0.0175	0.3139	0.6762
Standard deviation	0.5574	0.4575	0.0972
Kurtosis	1.7873	1.6746	3.0009
Skewness	0.5717	-0.2462	-0.9368
$p_{\text{JB}}$	0.0019	0.001	0.0011

correlation becomes negative again because of the second wave of European debt crisis. In general, we find that the correlations of Renminbi and Euro exhibit significant fluctuating behaviors over different periods. They are stable over relatively short periods of time and shift frequently between different regimes. This observation also implies that the time varying features of correlations are the results from the joint influences of common latent factors such as the global financial crisis, as well as the individual risk factor such as China's tight exchange rate bands.

**4.3. Experiment Results.** To evaluate the performance of the proposed algorithm against the benchmark ones, we use MSE to measure its predicative accuracy and Clark West test of equal predictive accuracy to test the statistical significance of the predictive accuracy [49, 50].

MSEs for the benchmark vector random walk (VRW) and VAR model are  $3.8774_{(\times 10^{-5})}$  and  $3.8737_{(\times 10^{-5})}$ . Then the wavelet denoising VAR model is applied to the testing data to investigate the effects of different parameters on the

model performance. Different combinations of parameters choices are pooled into the parameters pool, including two shrinkage rules (i.e., hard and soft shrinkage rules), decomposition levels up to scale 9, and 2 wavelet families including Daubechies 2 and Coiflet 2. The universal threshold is used during the denoising process. The model orders for VAR ( $r, m$ ) processes are determined following AIC and BIC minimization principle.

Experiment results in Table 4 further confirm that the forecasting accuracy of the proposed wavelet denoising model is sensitive to the wavelet parameters used to denoise the original data. Some of the wavelet parameters can improve the forecasting performance to the level that beats the traditional benchmark models significantly. For example, the wavelet denoising VAR outperforms both VRW and VAR model when Coiflet 2 wavelet is used for both hard and soft shrinkage rules.

Different criteria can then be used to determine the model specifications and parameters. Following the MSE minimization principle when identifying the appropriate model specifications, the chosen model specifications from results in Table 4 are Coiflet 2 wavelet family with hard threshold strategy at decomposition level 7.

Meanwhile, we propose the entropy maximization principle as the criteria when identifying the appropriate model specifications. Experiment results for different model specifications and parameters are listed in Table 5.

From results in Table 5, the chosen model specifications are Coiflet 2, with hard threshold strategy at decomposition level 3.

Experiment results in Table 6 show that the proposed algorithm outperforms the benchmark VRW and VAR model, in terms of predictive accuracy. The Clark West test of equal predictive accuracy suggests that the performance gap is significant at 7% confidence level against the VRW model and 13% confidence level against VAR model. Meanwhile, experiment results also show that adopting the proposed entropy measure during the model parameter determination would lead to lower MSE and higher  $P$  value for its statistical significance performance out-of-sample compared to the currently dominant MSE measure. Out-of-sample performance comparisons for the proposed MWVAR using different criteria during the model parameter determination in-sample suggest that using the proposed entropy measure leads to improved model performance out-of-sample. The

TABLE 4: MSE of MWVAR with different wavelet parameters and model specifications.

Wavelets	$Db2_H$ ( $\times 10^{-5}$ )	$Db2_S$ ( $\times 10^{-5}$ )	$Coiflet2_H$ ( $\times 10^{-5}$ )	$Coiflet2_S$ ( $\times 10^{-5}$ )
1	3.8863	3.8766	3.8660	3.8704
2	3.8757	3.8696	3.8524	3.8622
3	3.8771	3.8650	3.8529	3.8604
4	3.8814	3.8688	3.8531	3.8611
5	3.8816	3.8726	3.8524	3.8604
6	3.8817	3.8728	3.8528	3.8594
7	3.8818	3.8729	3.8517	3.8599
8	3.8819	3.8741	3.8519	3.8594
9	3.8818	3.8746	3.8519	3.8573
Average	<b>3.8810</b>	<b>3.8719</b>	<b>3.8539</b>	<b>3.8612</b>

TABLE 5: Entropy of MWVAR prediction with different wavelet parameters and model specifications.

Wavelets	$Db2_H$ ( $\times 10^{-5}$ )	$Db2_S$ ( $\times 10^{-5}$ )	$Coiflet2_H$ ( $\times 10^{-5}$ )	$Coiflet2_S$ ( $\times 10^{-5}$ )
1	3.3462	1.9278	0.5385	0.9922
2	3.7502	3.0319	2.1602	2.3601
3	3.9625	3.1196	2.5694	2.7095
4	3.8269	3.1158	2.5220	2.5955
5	3.8084	3.2815	2.4967	3.1538
6	3.8023	3.4627	2.4291	3.2342
7	3.8048	3.5441	2.5636	3.4439
8	3.8031	3.6604	2.5646	3.4929
9	3.7963	3.7501	2.5565	3.6166

TABLE 6: Performance evaluation based on MSE.

	MSE ( $\times 10^{-5}$ )	$P_{CW,RW}$	$P_{CW,VAR}$
<b>VRW</b>	2.9557	N/A	0.1017
<b>VAR</b>	2.9518	N/A	N/A
MWVAR <sub>MSE</sub>	2.9481	0.0796	0.1537
MWVAR <sub>Entropy</sub>	2.9472	0.0671	0.1222

proposed MWVAR algorithm has achieved the lower MSE in general, compared with the algorithm performance using in-sample MSE as the criteria.

The performance improvement is attributed to the analysis of latent risk structure in the multiscale domain using wavelet analysis, as well as the best basis and parameters selection based on entropy maximization principle. These results further imply that the exchange rate data is complicated processes with a mixture of underlying DGPs of different natures. There may be redundant representation of the underlying latent structure since there lacks explicit and analytic solutions. The determination of best basis and parameters, heterogeneous in nature, hold the key to further performance improvement and more thorough understanding of the DGPs during the modeling process.

## 5. Conclusions

In this paper, we propose the HMM-based theoretical framework for exchange rate forecasting. Under the proposed theoretical framework we further propose the multivariate wavelet-based exchange rate forecasting algorithm, as one particular implementation. We find that the exchange rate behaviors are affected by noises and main trends, which have different characteristics. The separation of noises and data need to be conducted in a multiscale manner to recover the useful data for further modeling by VAR model. Results from empirical studies using the USD against RMB and Euro, as the typical pair, confirm the performance improvement of the proposed models, against the benchmark models. Work done in this paper suggests that the more accurate separation of data and noises leads to better behaved data and higher level of model generalizability.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The authors would like to express their sincere appreciation to the editor and the three anonymous referees for their valuable comments and suggestions, which helped improve the quality of the paper tremendously. This work is supported by the CityU Strategic Research Grant (no. 7004135), the National Natural Science Foundation of China (NSFC nos. 71201054 and 91224001), and the Fundamental Research Funds for the Central Universities (no. ZZ1315).

## References

- [1] L. Sarno and G. Valente, "Empirical exchange rate models and currency risk: some evidence from density forecasts," *Journal of International Money and Finance*, vol. 24, no. 2, pp. 363–385, 2005.
- [2] A. A. Salisu and H. Mobolaji, "Modeling returns and volatility transmission between oil price and US-Nigeria exchange rate," *Energy Economics*, vol. 39, pp. 169–176, 2013.
- [3] W. Chkili and D. K. Nguyen, "Exchange rate movements and stock market returns in a regime-switching environment: evidence for BRICS countries," *Research in International Business and Finance*, vol. 31, pp. 46–56, 2014.
- [4] R. S. Hacker, H. K. Karlsson, and K. Mansson, "An investigation of the causal relations between exchange rates and interest rate differentials using wavelets," *International Review of Economics & Finance*, vol. 29, pp. 321–329, 2014.
- [5] B. H. Kim, H. Kim, and H. G. Min, "Reassessing the link between the Japanese yen and emerging Asian currencies," *Journal of International Money and Finance*, vol. 33, pp. 306–326, 2013.
- [6] R. A. Meese and K. Rogoff, "Empirical exchange rate models of the seventies: do they fit out of sample?" *Journal of International Economics*, vol. 14, no. 1-2, pp. 3–24, 1983.

- [7] F. Wen, Z. Li, C. Xie, and D. Shaw, "Study on the fractal and chaotic features of the Shanghai composite index," *Fractals-Complex Geometry Patterns and Scaling in Nature and Society*, vol. 20, no. 2, pp. 133–140, 2012.
- [8] F. Wen and X. Yang, "Skewness of return distribution and coefficient of risk premium," *Journal of Systems Science and Complexity*, vol. 22, no. 3, pp. 360–371, 2009.
- [9] W. Huang, K. K. Lai, Y. Nakamori, S. Wang, and L. Yu, "Neural networks in finance and economics forecasting," *International Journal of Information Technology & Decision Making*, vol. 6, no. 1, pp. 113–140, 2007.
- [10] L. Yu, S. Y. Wang, and K. K. Lai, "A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates," *Computers & Operations Research*, vol. 32, no. 10, pp. 2523–2541, 2005.
- [11] F. Wen and Z. Dai, "Modified Yabe-Takano nonlinear conjugate gradient method," *Pacific Journal of Optimization*, vol. 8, no. 2, pp. 347–360, 2012.
- [12] E. Peters, *Fractal Market Analysis : Applying Chaos Theory to Investment and Economics*, John Wiley & Sons, London, UK, 1994.
- [13] R. Gencay, F. Selcuk, and A. Ulugulyagci, "High volatility, thick tails and extreme value theory in value-at-risk estimation," *Insurance: Mathematics and Economics*, vol. 33, no. 2, pp. 337–356, 2003.
- [14] A. K. Tiwari, A. B. Dar, and N. Bhanja, "Oil price and exchange rates: a wavelet based analysis for India," *Economic Modeling*, vol. 31, pp. 414–422, 2013.
- [15] J. C. Reboredo and M. A. Rivera-Castro, "A wavelet decomposition approach to crude oil price and exchange rate dependence," *Economic Modeling*, vol. 32, pp. 42–57, 2013.
- [16] A. G. Orlov, "A cospectral analysis of exchange rate comovements during Asian financial crisis," *Journal of International Financial Markets, Institutions and Money*, vol. 19, no. 5, pp. 742–758, 2009.
- [17] S. Blanco, A. Figliola, R. Q. Quiroga, O. A. Rosso, and E. Serrano, "Time-frequency analysis of electroencephalogram series. III. Wavelet packets and information cost function," *Physical Review E*, vol. 57, no. 1, pp. 932–940, 1998.
- [18] R. R. Coifman and M. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 713–718, 1992.
- [19] P. Xu, X. Hu, and D. Yao, "Improved wavelet entropy calculation with window functions and its preliminary application to study intracranial pressure," *Computers in Biology and Medicine*, vol. 43, no. 5, pp. 425–433, 2013.
- [20] A. Samui and S. R. Samantaray, "Wavelet singular entropy-based islanding detection in distributed generation," *IEEE Transactions on Power Delivery*, vol. 28, no. 1, pp. 411–418, 2013.
- [21] D. Wang, D. Miao, and C. Xie, "Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection," *Expert Systems with Applications*, vol. 38, no. 11, pp. 14314–14320, 2011.
- [22] R. J. Bessa, V. Miranda, and J. Gama, "Entropy and correntropy against minimum square error in offline and online three-day ahead wind power forecasting," *IEEE Transactions on Power Systems*, vol. 24, no. 4, pp. 1657–1666, 2009.
- [23] D. Kwon, M. Vannucci, J. J. Song, J. Jeong, and R. M. Pfeiffer, "A novel wavelet-based thresholding method for the pre-processing of mass spectrometry data that accounts for heterogeneous noise," *Proteomics*, vol. 8, no. 15, pp. 3019–3029, 2008.
- [24] D. Boto-Giralda, F. J. Diaz-Pernas, D. Gonzalez-Ortega et al., "Wavelet-based denoising for traffic volume time series forecasting with self-organizing neural networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 25, no. 7, pp. 530–545, 2010.
- [25] J. Gao, H. Sultan, J. Hu, and W.-W. Tung, "Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: a comparison," *IEEE Signal Processing Letters*, vol. 17, no. 3, pp. 237–240, 2010.
- [26] U. Lotric and A. Dobnikar, "Predicting time series using neural networks with wavelet-based denoising layers," *Neural Computing & Applications*, vol. 14, no. 1, pp. 11–17, 2005.
- [27] U. Lotric, "Wavelet based denoising integrated into multilayered perceptron," *Neurocomputing*, vol. 62, no. 1–4, pp. 179–196, 2004.
- [28] B. Vidakovic, *Statistical Modeling by Wavelets*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, John Wiley & Sons, London, UK, 1999.
- [29] M. Jansen, *Second Generation Wavelets and Applications*, Springer, London, UK, 2005.
- [30] G. P. Nason, *Wavelet Methods in Statistics with R*, Springer, London, UK, 2008.
- [31] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, Pa, USA, 1992.
- [32] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.
- [33] D. Donoho and I. Johnstone, "Ideal denoising in an orthonormal basis chosen from a library of bases," *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique*, vol. 319, no. 12, pp. 1317–1322, 1994.
- [34] J. E. Bates and H. K. Shepard, "Measuring complexity using information fluctuation," *Physics Letters A*, vol. 172, no. 6, pp. 416–425, 1993.
- [35] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.
- [36] W. A. Brock and C. H. Hommes, "A rational route to randomness," *Econometrica*, vol. 65, no. 5, pp. 1059–1095, 1997.
- [37] W. A. Brock and A. W. Kleidon, "Periodic market closure and trading volume: a model of intraday bids and asks," *Journal of Economic Dynamics and Control*, vol. 16, no. 3–4, pp. 451–489, 1992.
- [38] W. A. Brock and B. D. LeBaron, "A dynamic structural model for stock return volatility and trading volume," *The Review of Economics and Statistics*, vol. 78, no. 1, pp. 94–110, 1996.
- [39] C. H. Hommes, "Financial markets as nonlinear adaptive evolutionary systems," *Quantitative Finance*, vol. 1, no. 1, pp. 149–167, 2001.
- [40] M. M. Dacorogna, R. Gençay, U. A. O. Muller, O. V. Pictet, and R. B. Olsen, *An Introduction to High-Frequency Finance*, Academic Press, San Diego, Calif, USA, 2001.
- [41] J. K. Goeree and C. Hommes, "Heterogeneous beliefs and the non-linear cobweb model," *Journal of Economic Dynamics and Control*, vol. 24, no. 5–7, pp. 761–798, 2000.
- [42] J. Farmer, "Market force, ecology and evolution," *Industrial and Corporate Change*, vol. 11, no. 59, pp. 895–953, 2002.
- [43] G. Athanasopoulos and F. Vahid, "VARMA versus VAR for macroeconomic forecasting," *Journal of Business & Economic Statistics*, vol. 26, no. 2, pp. 237–252, 2008.



- [44] S. T. Rachev, S. Mittnik, F. J. Fabozzi, S. M. Focardi, and T. Jašić, Eds., *Financial Econometrics: From Basics to Advanced Modeling Techniques*, John Wiley & Sons, Chichester, UK, 2007.
- [45] P. R. Hansen and A. Timmermann, "Choice of sample split in out-of-sample forecast evaluation," *Economics Working Papers ECO2012/10*, European University Institute, 2012.
- [46] H. Zou, G. Xia, F. Yang, and H. Wang, "An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting," *Neurocomputing*, vol. 70, no. 16–18, pp. 2913–2923, 2007.
- [47] C. M. Jarque and A. K. Bera, "A test for normality of observations and regression residuals," *International Statistical Review*, vol. 55, no. 2, pp. 163–172, 1987.
- [48] W. A. Broock, J. A. Scheinkman, W. D. Dechert, and B. LeBaron, "A test for independence based on the correlation dimension," *Econometric Reviews*, vol. 15, no. 3, pp. 197–235, 1996.
- [49] A. Alves Portela Santos, N. Carneiro Affonso da Costa Jr., and L. dos Santos Coelho, "Computational intelligence approaches and linear models in case studies of forecasting exchange rates," *Expert Systems with Applications*, vol. 33, no. 4, pp. 816–823, 2007.
- [50] R. S. Mariano and F. X. Diebold, "Comparing predictive accuracy," *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 253–263, 1995.

