Proceedings of DET2011
7th International Conference on Digital Enterprise Technology
Athens, Greece
28-30 September 2011

DYNAMIC WAVELET NEURAL NETWORK MODEL FOR FORECASTING RETURNS OF SHFE COPPER FUTURES PRICE

Li Shi
Dept. of IMSE, The University of Hong Kong
amyshi0629@hku.hk

LK Chu
Dept. of IMSE, The University of Hong Kong
lkchu@hkucc.hku.hk

Yuhua Chen
Research institute of Tsinghua University in Shenzhen
jasonchen1227@gmail.com

ABSTRACT

Appropriate forecasting of commodity futures price returns is of crucial importance to achieve hedging effectiveness against the returns volatility risk. This paper presents a nonparametric dynamic recurrent wavelet neural network model for forecasting returns of Shanghai Futures Exchange (SHFE) copper futures price. The proposed model employs a wavelet basis function as the activation function for hidden-layer neurons of the neural network. The aim of this arrangement is to incorporate the fractal properties discovered in futures price return series. In the wavelet transform domain, fractal self-similarity information of the returns series over a certain time scale can be extracted. Input variables are analyzed and selected to facilitate effective forecasting. Statistical indices such as normal mean square error (NMSE) are adopted to evaluate forecasting performance of the proposed model. The forecasted result shows that dynamic wavelet neural network has good prediction properties compared with traditional linear statistical model such as ARIMA and other neural network forecasting models.

KEYWORDS

Wavelet Neural Networks, SHFE Copper Futures, Forecasting, Financial Time Series, Fractal Market

1. INTRODUCTION

Copper is an important material for industrial production and yet its volatile price movement has been a major concern of the manufacturing industry in the past few years. Some practitioners choose to use copper's exchange-traded commodity futures for hedging against such price volatility. The availability of an effective approach for forecasting copper futures price is crucial in the performance of such hedging in order to optimize production plans or investment portfolios.

Since China is the world's largest consumer of copper, Shanghai Futures Exchange (SHFE) has hosted active trading of copper futures contracts and become an important venue for deciding the market price of copper price. Therefore, the study of the price fluctuation of SHFE copper futures has assumed importance as it is a prerequisite for effective forecasting. This study develops a dynamic recurrent wavelet neural network model for forecasting the returns of SHFE copper futures prices. The model will try to minimise forecasting errors and enhance forecasting capability compared with other approaches.

The popular view of commodity futures prices is due the theory of storage originating in the work of Kaldor (1939), Working (1948), and Brennan (1958). They try to explain the difference between synchronous spot and futures commodity prices in terms of storage carrying cost and convenience yield on inventory. The theory is further developed by Gibson and Schwartz (1990a), Miltersen and Schwartz (1998), and Schwartz (1997), etc. into

term structure commodity futures price models. These models are based on storage theory. However, they do not seem to possess the flexibility in explaining futures price fluctuation.

Since the simplest random walk model is proved to perform better than many complex structural models in financial time series forecasting (Meese and Rogoff, 1983, 1986), researchers have started using time series analysis to forecast commodity futures price (Taylor, 1986). Numerous studies have found that univariate time series such as Box-Jenkins ARIMA model performs well in forecasting commodity futures prices (Lawera, 2000). Time series analysis is useful in forecasting but lacks an economic foundation. In this paper, an ARIMA model of copper futures price will be applied as a benchmark for comparing with the proposed model.

Commodity futures prices are not only determined by supply and demand but are also susceptible to inter-market capital movements, political developments and so on. These factors contribute to strongly fluctuating and non-linear price behaviours. Since neural networks have a high level of nonlinear approximation and adaptive self-learning capabilities, they offer enormous potential for the construction of a nonlinear forecasting model of commodity futures prices based on certain facts dataset. The use of neural network for forecasting various commodity or commodity futures prices has been extensively studied, for example, by Grudnitski and Osburn (1993), and Zou etc., (2007). However, most existing studies on copper futures prices have used simple Back-Propagation neural networks (BPNN) in their forecasting models. A novel recurrent neural network forecasting model is proposed in this paper. It has input-output feedback and hidden neuron self-feedback incorporated into BPNN. These feedback loops create internal states of the network which allows it to exhibit dynamic temporal behaviour, and hence enhance the nonlinear approximation ability.

An important feature of the forecasting approach proposed in this paper lies in the use of wavelets as the activation function for feature detection by the neural network. Wavelet can be used to approximate target functions with good effect. The combination of wavelets and neural network can form a powerful forecasting model. Wavelets were first integrated with neural network by Zhang and Benveniste (1992), and various architectures are then developed by Pati and Krishnaprasad (1993), Zhang, J. (1995), Zhang (1997), Billings (2005), etc. Commodity futures prices have been proved to possess self-similarity and self-affinity structure, based on the fractal market hypothesis and fractal theory (Mandelbrot, 1982; Edgar, 1996). Since wavelet analysis procedure is implemented with temporal translation and dilation of a mother wavelet, they are found to be powerful in approximating commodity prices. The proposed model employs wavelet basis function as the activation function of hidden-layer, whose purpose is to incorporate the fractal properties of SHFE copper futures price.

This paper is organized as follows. In Section 2, input data pre-processing and selection is discussed for constructing a forecasting model expression. In Section 3, the basics of wavelet transform are first described. Then, a proper wavelet function is chosen for better fitting to the target time series. This is followed by the description of the architecture and training of proposed model. Section 4 gives the forecasting results of SHFE copper futures price returns using proposed model. The concluding remarks are given in Section 5.

2. FORECASTING MODEL EXPRESSION AND INPUT DATA SELECTION

As mentioned in Section 1, there exist two kinds of data that can be used to construct a forecasting model of copper futures prices. One is the external data, which include copper futures storage level and inter-market influence. The other is copper futures price historical data. As such, the nonlinear forecasting model is constructed as follows:

$$p(t) = f(p_{t-1}, p_{t-2}, ..., p_{t-l}) + g(x_{t-1}^1, x_{t-1}^2, ..., x_{t-1}^m) + e(t) \quad (1)$$

where $p(t)$ is the copper futures price series; $(p_{t-1}, p_{t-2}, ..., p_{t-n})$ the historical data of the $p(t)$; $f(.)$ the representation of a nonlinear auto-regression function; $(x_{t-1}^1, x_{t-1}^2, ..., x_{t-1}^m)$ the external factors; $g(.)$ a nonlinear mapping function and $e(t)$ the error between the forecasted and real prices.

It is obvious that the above dataset are non-stationary time series, which have to be scaled in order to have their non-stationary components removed. Such scaling is required for reducing the search space of the neural networks, and help to obtain the optimal coefficient easily. Logarithmic first difference is applied to transform the non-stationary time series into a stationary series (McNelis, 2005). Considering that prices change little or even do not change between two neighbouring days, the logarithmic first difference of every five days is taken. The following scaling function is applied:

$$\Delta \log p_t = \log p_t - \log p_{t-5} = \log(1 + \frac{p_t - p_{t-5}}{p_{t-5}}) \approx \frac{p_t - p_{t-5}}{p_{t-5}} \quad (2)$$

So, $\Delta \log p_t$ has the meaning of weekly price return rate. This approach will forecast the weekly return rate of copper futures price instead of the prices.

Actually, after the forecasted value of the rate is determined, the copper futures price can be readily obtained by applying the scaling function. Then equation (1) can be scaled and represented as follows:

$$r(t) = f(r_{t-1}, r_{t-2}, ..., r_{t-l}) + g(r(x_{t-1}^1), r(x_{t-1}^2), ..., r(x_{t-1}^m)) + e(t) \quad (3)$$

Five most influential external factors are selected to form the input vectors $(x_{t-1}^1, x_{t-1}^2, ..., x_{t-1}^m)$. First, Changjiang copper spot price in China and copper futures inventory level in the SHFE appointed warehouse are chosen based on the traditional storage price theory. They are denoted by SP and INV respectively. Then, London Metal Exchange (LME) three-month copper price is chosen since LME and SHFE copper futures prices have been shown to exhibit significant correlation (Zhang, 2003). WTI crude oil price is also included because crude oil can impact economy welfare and indirectly influence the copper futures price. Since crude oil and other energy sources are priced in US dollar, EUR/USD exchange rate is also adopted as an input, which is denoted by EUR. The input and output variables are listed in table 1.

Table 1 – Input and output variables

	Denotation	Equations
Input	$r(P-l)$ $l = 1.....l;$	$100 \cdot (\ln p_{t-5\cdot(l-1)} - \ln p_{t-5\cdot l})$
	r(INV)	$100 \cdot (\ln inv_t - \ln inv_{t-5})$
	r(SP)	$100 \cdot (\ln sp_t - \ln sp_{t-5})$
	r(LME)	$100 \cdot (\ln lme_t - \ln lme_{t-5})$
	r(WTI)	$100 \cdot (\ln wti_t - \ln wti_{t-5})$
	r(EUR)	$100 \cdot (\ln eur_t - \ln eur_{t-5})$
Output	r(P)	$100 \cdot (\ln p_{t+5} - \ln p_t)$

3. DYNAMIC WAVELET NEURAL NETWORKS

3.1. WAVELET TRANSFORM

This section will briefly describe wavelet transform, and discuss how wavelets can be applied to reconstruct functions or data series. Wavelet transform can be divided into two categories, which include continuous wavelet transform (CWT) and discrete wavelet transform (DWT).

The continuous wavelet transform $CWT_f(a,\tau)$ of function $f(t)$ is given by Daubechies, (1992):

$$CWT_f = \langle f(t), \psi_{a,\tau}(t) \rangle = \frac{1}{\sqrt{a}} \int_{\Re} f(t) \cdot \psi^* \left(\frac{t-\tau}{a} \right) dt \quad (4)$$

where $\psi(t)$ is the mother wavelet function, and $\psi(t) \in L^2(\Re)$. Its Fourier transform has to satisfy the following condition:

$$C_\psi = \int_{\Re} \frac{|\hat{\psi}(\omega)|}{|\omega|} d\omega < \infty \quad (5)$$

a and τ are dilation and translation parameter respectively. In CWT, a and τ vary continuously over $\Re$ (with the constraint $a \neq 0$). By scaling and shifting the mother wavelet, a set of wavelet basis functions are obtained as follows:

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \psi \left(\frac{t-\tau}{a} \right); \quad a, \tau \in \Re \quad (6)$$

Equation (4) transforms $f(t)$ from the time domain into the wavelet domain (a, τ domain). In the wavelet domain, frequency information at certain time can be obtained. It means that in the time domain, wavelet basis functions can be used to approximate both the smooth global variation and sharp local variation of the function. $f(t)$ can be reconstructed from wavelet basis functions by using an inverse wavelet transform, which is given as follows:

$$f(t) = \frac{1}{C_\psi} \int_0^{+\infty} \frac{da}{a^2} \int_{-\infty}^{+\infty} CWT_f(a,\tau) \cdot \frac{1}{\sqrt{a}} \psi \left(\frac{t-\tau}{a} \right) d\tau \quad (7)$$

In DWT, both a and τ take discrete values only. For binary representation, it is convenient to sample a and τ based on the so called "dyadic" grid. This special case of DWT is defined as dyadic wavelet transform. In this case, a and τ are represented as:

$$a_j = 2^{-j}, \quad \tau_{j,k} = k \cdot 2^{-j}; \quad j, k \in Z \quad (8)$$

Thus the definition of dyadic discrete wavelet is

$$\psi_{j,k}(t) = 2^{\frac{-j}{2}} \psi \left(2^{-j} \cdot t - k \right); \quad j, k \in Z \quad (9)$$

where j represents the number of wavelet basis functions, and k determines the time position of the wavelets.

Both CWT and DWT can be incorporated into neural networks as activation function. The existing wavelet neural networks can be categorized into two types (Billings, 2005). The adaptive wavelet neural network has wavelets as the activation function, which is obtained by performing CWT. The unknown parameters of such network include the weighting coefficients of the network and the dilation and translation factors of the wavelets. The other type is fixed grid wavelet neural network, whose activation function is obtained from the DWT. In such a wavelet neural network, the position and dilations of wavelets are predetermined,

and only the weighting coefficients need to be optimized.

In this paper, an adaptive wavelet neural network is developed to achieve the desired flexibility and for the accurate reconstruction of continuous time series. For practical implementation and computational efficiency, the inverse wavelet transform (equation (4)) can be expressed as

$$f(t)=\sum_{j,k} w_{j,k}\cdot\psi_{j,k}(t) \qquad (10)$$

Equation (10) is used as the expression of nonlinear mapping function $f(r_{t-1},r_{t-2},...,r_{t-n})$ in equation (3). So, the model can track the self-similarity and self-affinity properties of the return series, and result in a better approximation.

3.2. WAVELET FUNCTION SELECTION

An appropriate wavelet has to be selected in order to better reconstruct $f(t)$. Since wavelets will be employed as activation function in neural networks, the wavelet functions have to be differentiable and the Mexican hat wavelet (See figure 1) function is chosen. This wavelet function is commonly used in time series decomposition and reconstruction. Its function is expressed as follows:

$$\psi(t)=Ce^{-\frac{t^2}{2}}\cos(5t) \qquad (11)$$

Mexican hat wavelet offers other advantages for reconstructing $f(t)$. First, Mexican hat wavelet is based on CWT, is symmetrical, and provides an exact time frequency analysis. This makes it a good choice to process data that vary continuously in time. Second, Mexican hat wavelet has a rapid vanishing function, which leads to an accurate and efficient approximation of target time series.

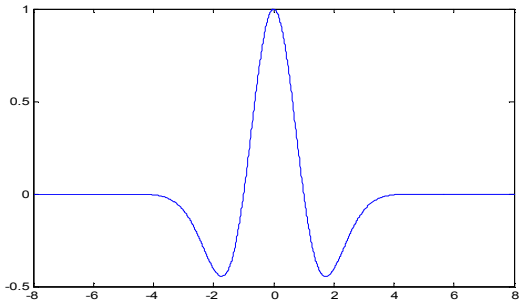

Figure 1 – Mexican hat wavelet

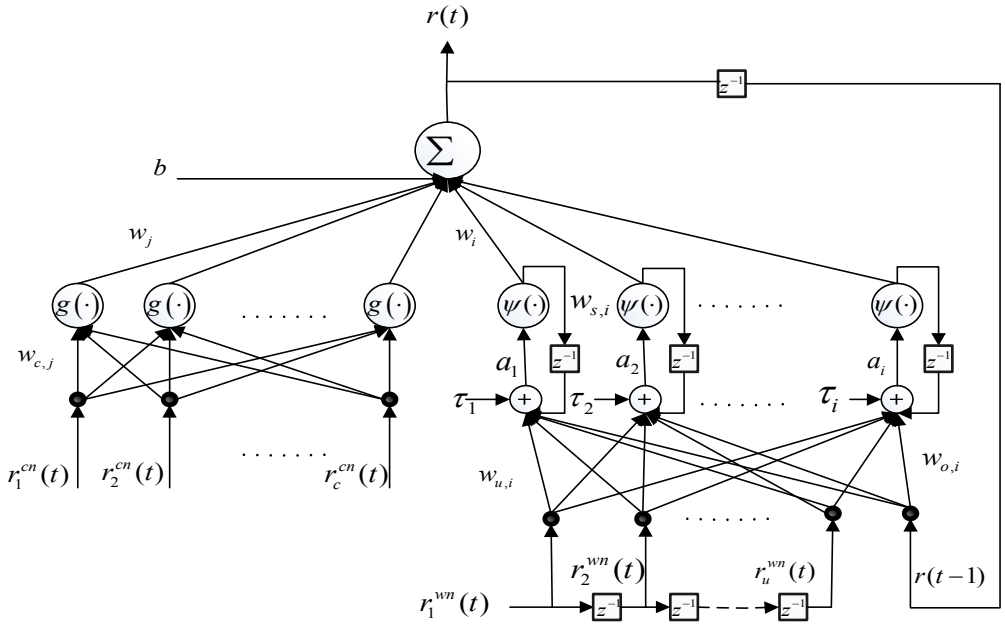

Figure 2 – Structure of proposed recurrent wavelet neural network

3.3 ARCHITECTURES OF A RECURRENT WAVELET NEURAL NETWORK (RWNN)

As mentioned above, the returns of SHFE copper futures price originate from the storage theory, and are inevitably influenced by inter-market factors. In order to extract useful information from related factors for effective forecasting, commonly used sigmoid functions are adopted to produce a nonlinear mapping from these databases to target returns. The sigmoid active function used in the conventional neural network part of the model is expressed as follows:

$$g(x_j(t))=1/\left[1+\exp(-x_j(t))\right]$$

$$x_j(t)=\sum_{c=1}^{5} w_{c,j}\cdot r_c^{cn}(t)$$

$$r_c^{cn}=r(INV),r(SP),r(LME),r(WTI),r(EUR) \qquad (12)$$

where x_j is the input of hidden neuron j in the conventional neural network part. r_c^{cn} is the input of the conventional neural network part as listed in table 1.

In order to incorporate dynamic temporal behaviour and enhance the nonlinear approximate ability, a dynamic wavelet neural network with feedback topology is developed. In this model, the feed-forward part consists of a wavelet network combined with a conventional neural network using sigmoid activation function. The feedback part includes input-output feedback loop and wavelet neuron self-feedback loop. Therefore, the nonlinear estimator in the proposed forecasting model can be expressed as follows:

$$r(t)=\sum_{i=1}^{M} w_i\cdot\psi_i\left(\frac{y_i(t)-\tau_i}{a_i}\right)+\sum_{j=1}^{N} w_j\cdot g(x_j(t))+b$$

$$y_i(t)=\sum_{u=1}^{l} w_{u,i}\cdot r_u^{wn}(t)+w_{o,i}\cdot r(t-1)+w_{s,i}\cdot y_i(t-1)$$

$$r_u^{wn}(t)=r\left(P-1\right),.....r\left(P-l\right) \qquad (13)$$

where y_i is the input of hidden neuron j of the wavelet neural network part. r_u^{wn} is the input of the conventional neural network part, listing before in table 1.

The architecture of the proposed dynamic recurrent neural network model is shown in figure 2.

3.4 TRAINING THE PROPOSED NEURAL NETWORK

In order to obtain all the parameters in the model, i.e. weights, bias, and translation and dilation parameters, a learning algorithm is used for training the network. In this paper, an improved real-time training algorithm for recurrent networks called real-time recurrent learning algorithm (RTRL) (Williams e.a., 1995) is adopted due to of its fast convergence for an accurate training.

RTRL is a real-time back-propagation (BP) gradient descent training algorithm. It does not employ the error measure, which is obtained by summing up the error between real returns and model output (equation (14)) during the training period. Instead, only the instantaneous error measure (see equation (15)) is used for calculating parameter updates at each time instant of the continually running network.

$$E_{Total}=\frac{1}{2}\sum_{t=t_0}^{t_n}\left(r_{real}(t)-r(t)\right)^2 \qquad (14)$$

$$E(t)=\frac{1}{2}\left(r_{real}(t)-r(t)\right)^2 \qquad (15)$$

So, at every time t, the parameters are adapted according to

$$\Delta par.(t)=-\eta\cdot\frac{\partial E(t)}{\partial par.} \qquad (16)$$

where η is the learning rate. Thus in this model, the added value of each parameter is calculated in the following equations:

$$\Delta b=-\eta\cdot\frac{\partial E(t)}{\partial r(t)}\cdot\frac{\partial r(t)}{\partial b}=\eta\cdot(r_{real}(t)-r(t)) \qquad (17)$$

$$\Delta w_j=-\eta\cdot\frac{\partial E(t)}{\partial r(t)}\cdot\frac{\partial r(t)}{\partial w_j}=\eta\cdot(r_{real}(t)-r(t))\cdot g_j(t) \qquad (18)$$

$$\Delta w_i=-\eta\cdot\frac{\partial E(t)}{\partial r(t)}\cdot\frac{\partial r(t)}{\partial w_i}=\eta\cdot(r_{real}(t)-r(t))\cdot\psi_i(t) \qquad (19)$$

$$\Delta w_{c,j}=-\eta\cdot\frac{\partial E(t)}{\partial r(t)}\cdot\frac{\partial r(t)}{\partial g_j(t)}\cdot\frac{\partial g_j(t)}{\partial w_{c,j}}=\eta\cdot(r_{real}(t)-r(t))\cdot w_j\cdot g_j'(t)\cdot r_c^{cn}(t) \qquad (20)$$

$$\Delta w_{x,i}\left(w_{u,i}w_{s,i},w_{o,i}\right)=-\eta\cdot\frac{\partial E(t)}{\partial r(t)}\cdot\frac{\partial r(t)}{\partial\psi_i(t)}\cdot\frac{\partial\psi_i(t)}{\partial w_{x,i}}$$

$$=\eta\cdot(r_{real}(t)-r(t))\cdot w_i\cdot\psi_i'(t)\cdot(w_{s,i}\cdot\frac{\partial\psi_i(t-1)}{\partial w_{x,i}}+$$

$$w_{o.i}\cdot\frac{\partial r(t-1)}{\partial w_{x,i}}+\delta_{x,u}\cdot r_u^{wn}(t)+\delta_{x,s}\cdot\psi_i(t-1)+\delta_{x,o}\cdot r(t-1))$$

$$\frac{\partial\psi_i(t)}{\partial w_{x,i}}=\psi_i'(t)\cdot(w_{s,i}\cdot\frac{\partial\psi_i(t-1)}{\partial w_{x,i}}+w_{o.i}\cdot\frac{\partial r(t-1)}{\partial w_{x,i}}+\delta_{x,u}\cdot r_u^{wn}(t)+$$

$$\delta_{x,s}\cdot\psi_i(t-1)+\delta_{x,o}\cdot r(t-1))$$

$$\frac{\partial r(t)}{\partial w_{x,i}}=w_i\cdot\psi_i'(t)\cdot(w_{s,i}\cdot\frac{\partial\psi_i(t-1)}{\partial w_{x,i}}+w_{o.i}\cdot\frac{\partial r(t-1)}{\partial w_{x,i}}+\delta_{x,u}\cdot r_u^{wn}(t)+$$

$$+\delta_{x,s}\cdot\psi_i(t-1)+\delta_{x,o}\cdot r(t-1))$$

$$\frac{\partial\psi_i(0)}{\partial w_{x,i}}=0;\quad\frac{\partial r(0)}{\partial w_{x,i}}=0 \qquad (21)$$

$$\Delta a=-\eta\cdot\frac{\partial E(t)}{\partial r(t)}\cdot\frac{\partial r(t)}{\partial\psi_i(t)}\cdot\frac{\partial\psi_i(t)}{\partial a}=-\eta\cdot(r_{real}(t)-r(t))\cdot w_i\cdot\psi_i'(t)/a^2 \qquad (22)$$

$$\Delta\tau=-\eta\cdot\frac{\partial E(t)}{\partial r(t)}\cdot\frac{\partial r(t)}{\partial\psi_i(t)}\cdot\frac{\partial\psi_i(t)}{\partial\tau}=-\eta\cdot(r_{real}(t)-r(t))\cdot w_i\cdot\psi_i'(t) \qquad (23)$$

In order to prevent the RTRL process from being trapped in a local maximization, momentum factors are introduced for adjusting the network parameters. Taking w_j for brief description, the momentum factor of w_j can be expressed as:

$$m_{w_j}=k\cdot(w_j(t)-w_j(t-1)) \qquad (24)$$

Then the w_j(t+1) is updated in accordance with the following rules:

$$w_j(t+1) = w_j(t) + \Delta w_j(t+1) + k \cdot m_{w_j} \quad (25)$$

4. PERFORMANCE EVLUATION OF DYNAMIC RECURRENT WAVELET NEURAL NETWORK MODEL

4.1. DATA SETS

The input and output data are obtained and listed in table 1 for model evaluation.

First, the daily close price of copper futures and the INV are collected from the SHFE official website. Unlike LME, SHFE have the copper futures contracts expired in each month of the year. The exact three-month-to-maturity contracts may not be available on a given day. Since every trading day has different futures prices, in order to deal with the discontinuity of futures prices, the nearest to the three-month-to-maturity contract is used. The nearest to the three-month-to-maturity contract is selected for its actively trading. Since the first delivery day of a contract is the first business day of each month, at the beginning of each month, a nearest to the three-month-to maturity contract is selected and kept for a month. Then, at the beginning of the next month, the following nearest to the three-month-to-maturity contract will be selected for replacing the previous one. In this way, a series of continuous copper futures price can be formed.

Second, the daily close prices of LME, SP, WTI and EUR are collected from the website of Bloomberg. Third, the weekly returns series of the above obtained price are computed using the equations listing in table 1.

The sampling time window is from January 4th, 2005 to November 9th, 2010. A total of 1,400 samples are obtained. There exists significantly rise and fall economic cycle during this period, which make the proposed forecasting approach more comprehensive and meaningful. The weekly returns of SHFE copper futures price during this time window is shown in figure 3.

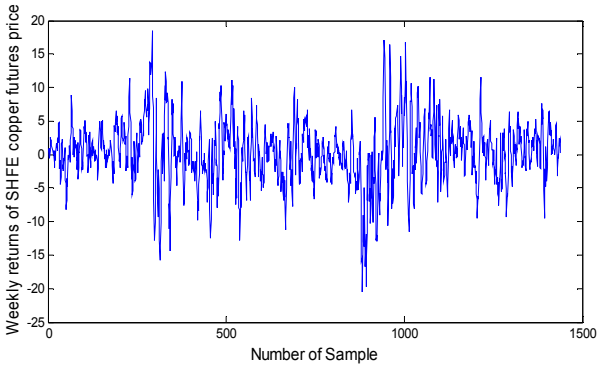

Figure 3 –Weekly returns of SHFE copper futures price

4.2 FINDING THE INPUT DIMENSION

The sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) are employed to determine the input dimension of the auto-regression part of the model. Figure 4 is plotted using weekly return rate in figure 3. The input dimension is found to be four.

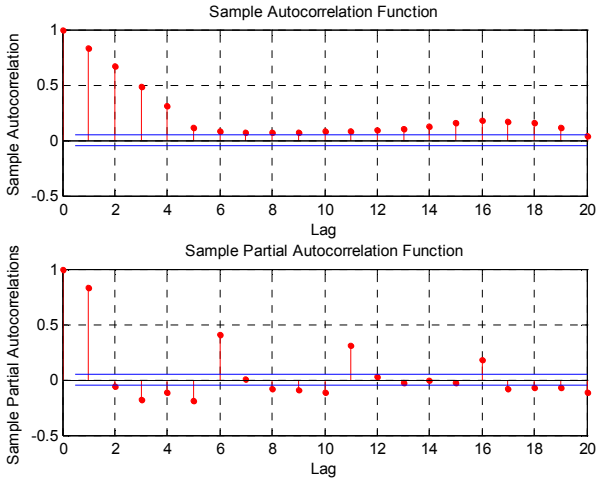

Figure 4 – ACF and PACF of copper futures price returns

4.3 FORECASTING ACCURATE ASSESSMENT

Several error measurements have been chosen to evaluate the performance of proposed model. Normalized mean square error (NMSE) and mean absolute error (MAE) are adopted to evaluate the error between target returns and the model output. Direction sign (DS) is adopted to show the hit ratio, defined by the model's output moving in the same direction as target returns. Table 2 lists above performance measurements and their equations.

Table 2 – Performance measurements

Measurements	Equations
NMSE	$\frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\frac{1}{n-1}\cdot\sum_{i=1}^{n}(y_i-\bar{y})^2}\ ;\bar{y}=\frac{1}{n}\cdot\sum_{i=1}^{n}y_i$
MAE	$\frac{1}{n}\cdot\sum_{i=1}^{n}\lvert y_i-\hat{y}_i\rvert$
DS	$\frac{100}{n}\cdot\sum_{i=1}^{n}d_i;\ d_i=\begin{cases}1 & (y_i-y_{i-1})(\hat{y}_i-\hat{y}_{i-1})\geq 0\\ 0 & otherwise\end{cases}$

4.4 FORECASTING RESULTS

In this approach, a fixed forecasting scheme is adopted. The scheme involves training and estimating parameters of the proposed model on the first 1,000 data in the dataset, and uses these estimates to produce all the forecasts for the

following out-of-sample data. Figure 5 shows the comparisons of the following 200 out-of-sample testing data and the estimated outputs using the proposed model. Figure 5(a) shows the forecasting targets and Figure 5(b) shows the absolute forecasting error.

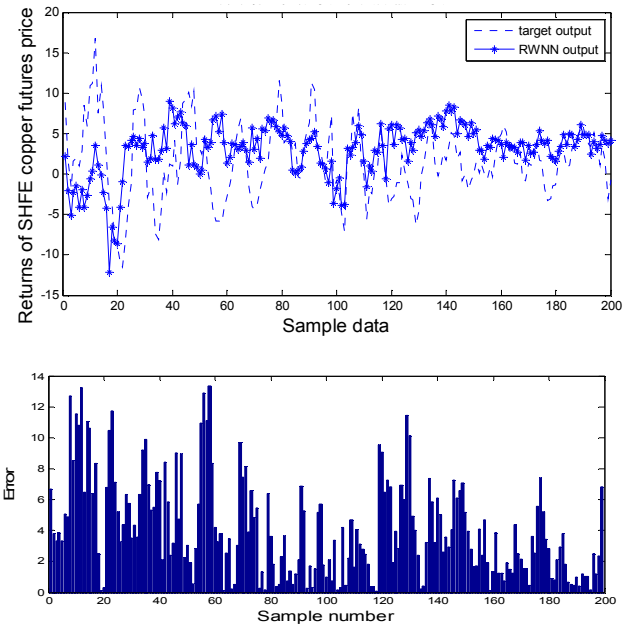

Figure 4 – comparison of out-of-sample and estimation data

Three different out-of-sample lengths (listing in table 3) are chosen for testing the forecasting performance of the proposed model. Variance testing data lengths are applied for testing the near-term and long-term forecasting ability of the proposed model.

Table 3 – In-sample and out-of-sample time windows

Sample	Time windows
In sample data	Feb. 1st, 2005 to Aug 18th, 2009 (1000 data)
Out of sample data	Aug. 19th, 2009 to Aug. 11st, 2009 (100 data)
	Aug. 19th, 2009 to Jan. 7th, 2010 (200 data)
	Aug. 19th, 2009 to Nov. 9th, 2010 (400 data)

Three other conventional forecasting models are also set up for comparing with our proposed model. The three models are feed-forward wavelet neural network model (WNN), fully recurrent BP neural network (FRNN), and ARIMA model. In figure 4, ACF is trailed and PACF truncated at lag 1. It indicates the need for an AR (1) model out of ARIMA family models. The results due to these comparisons are listing in table 4, 5 and 6.

Table 4 – Performance comparison when forecasting 100 data

	RWNN	WNN	FRNN	AR(1)
NMSE	0.8993	1.1070	1.3425	1.6657
MAE	4.3104	4.8678	5.0864	5.9252
DS	66	59	58	56

Table 5 – Performance comparison when forecasting 200 data

	RWNN	WNN	FRNN	AR(1)
NMSE	0.9593	1.3290	1.4327	1.8532
MAE	3.6325	4.5435	4.6325	7.2353
DS	63	55	56	54

Table 6 – Performance comparison when forecasting 400 data

	RWNN	WNN	FRNN	AR(1)
NMSE	1.0625	1.5242	1.5076	1.9543
MAE	3.6982	5.0634	4.8735	7.8963
DS	62	54	56	53

The forecasting results in table 4, 5 and 6 show that the dynamic recurrent wavelet model outperforms other neural network and ARIMA model both in value accuracy and directional accuracy. The model achieves the best performance in the near- and long-term forecasting.

5. CONCLUSIONS

This research has advanced the study of conventional neural network in forecasting returns of SHFE copper futures price by presenting a dynamic recurrent neural network model. The proposed model combines the feature detection property of wavelet and temporal memory behaviour of recurrent neural network for capturing the dynamics of copper futures returns and results in a better forecasting. The proposed model also considers the influence of exogenous factors and extracts useful information for assistant forecasting. The forecasting results show that the dynamic recurrent wavelet neural network model outperformed in the near- and long-term forecasting of returns of SHFE copper futures price compared to other conventional models.

6. ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments.

REFERENCES

Billings, Stephen A. and Wei, Hua-Liang, "A new class of wavelet networks for nonlinear system identification", *IEEE Transation on Neural networks,* Vol. 16, No. 4, 2005

Brennan, M. J., "The supply of storage", *American Economic Review*, Vol. 48, 1958, pp 50- 72

Daubechies, Ingrid, "Ten Lectures on Wavelets", Society for Industrial and Applied Mathematics Philadelphia, PA, USA, 1992

Edgar E. Peters, "Chaos and Order in the Capital Markets", John Wiley & Sons, NY. 1996

Gibson, R., and E. S. Schwartz. "Valuation of Long Term Oil-Linked Assets", In "Stochastic Models and Options Values", D. Lund and B. Oksental, Elsevier, Amsterdam, 1990

Grudnitski, Gary and Osburn, Larry, "Forecasting S&P and gold futures prices: An application of neural networks", *The Journal of Futures Markets*, Vol. 13, Issue 6, 1993, pp 631–643

Kaldor, N., "Speculation and economic stability", *Review of Economic Studies* Vol. 7, 1939, pp 1-27

Lawera, Martin Lukas, "Futures prices: Data mining and modelling approaches", *PHD thesis*

Mandelbrot B. B, "The fractal geometry of nature", W. H. Freeman, NY. 1982

Mcnelis, Paul D., "Neural neworks in finance, gaining predictive edge in the market". Elsevier Inc., Oxford, 2005

Meese, RA and Rogoff, K., "Empirical exchange rate models of the seventies: do they fit out-of-sample?" *Journal of international Economics.* Vol. 14, 1983, pp 3-24

Meese, RA and Rogoff, K. (1986). "Was it real? The exchange rate-interest differential relation over the modern floating-rate period", *Jounal of Finance*, Vol. 43, No. 4, 1983, pp 933-948

Miltersen, K. R., and E. Schwartz. "Pricing of Options on Commodity Futures with Stochastic Term Structure of Convenience Yields and Interest Rates", *Journal of Financial and Quantitative Analysis*, Vol. 33, 1998, pp 33-59

Nowrouz Kohzadia, Milton S. Boydb, Bahman Kermanshahib and Iebeling Kaastrac, "A comparison of artificial neural network and time series models for forecasting commodity prices", *Neurocomputing,* Vol. 10, Issue 2, 1996, pp 169-181

Pati, Y. C. and Krishnaprasad, P. S., "Analysis and synthesis of feed forward neural networks using dscrete affine wavelet transformations", *IEEE Transation on Neural Netwoks,* Vol. 4, No. 1, 1993

Schwartz, E. S. "The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging" *Journal of Finance,* Vol. 52, 1997, pp 923-973

Taylor, Stephen J., "Modelling financial time series", World Scientific, Singapore, 1986

Williams, R. J, Zipser, D., "Gradient-Based Learning Algorithms for Recurrent Networks and their computational complexity", Book chapter of "Backpropagation: Theory, Architectures and Applications", Lawrence Erlbaum Associates Pubulishers, New Jersey, USA

Working, H., "Theory of the inverse carrying charge in futures markets", *Journal of Farm Economics*, Vol. 30, 1948, pp 1-28

Zhang Jun, "Wavelet neural networks for function learning", *IEEE Transation on Signal Processing*, Vol. 43, No. 6, 1995

Zhang, Q, "Using wavelet network in nonparametric estimate", *IEEE Transation on Neural Netwoks,* Vol. 8, No. 2, 1997

Zhang, Q., and Benveniste, A., "Wavelet networks" *IEEE Transation on Neural Networks*, Vol. 3, No. 6, 1992

Zhang, G.P., "Correlation analysis between Shanghai copper futures and international copper futures markets", working paper, Shanghai Futures Exchange, 2003

Zou, H.F, Xia, G.P, Yang F.T. and Wang H. Y., "An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting", *Neurocomputing,* Vol. 70, Issue 16-18, 2007, pp 2913-2923