Wavelet and Hilbert-Huang Transform Based on Predicting Stock Forecasting in Second-Order Autoregressive Mode

Hong-Ru Ke, Kai-Cheng Wang, Chi-I. Yang, and Kuei-Fang Chang

Abstract—In this dissertation the Discrete Wavelet Transform and HHT Transform will be used for pre-treatment of the stock prices of 50 indicative companies in Taiwan 50 Index in order to establish the stationary state time series, and then the AR(2) model will be used for daily prediction of data in the next transaction days. In the end the root mean square error (RMSE) will be used for accuracy comparison, and the K-S test will be used to examine whether or not the predication residual is in normal distribution. The result of prediction has indicated that DWT prediction performs better than HHT prediction in the first week with a share of 31/50, yet HHT prediction performs better than DWT prediction in the second week with a share of 29/50. As for the two-week prediction, the performances of DWT and HHT predictions are similar. The true value will fall within the Wavelet/HHT confidence bands established by DWT or HHT method in the first week, the second week, or during the two-week time span. The scope of occurrence of this regression line can be obtained using the confidence band of this regression line, and the Wavelet/HHT confidence bands of the 50 underlying stocks established by DWT or HHT will all fall within this confidence band.

Index Terms—Discrete wavelet transform, HHT, second-order autoregressive model, wavelet confidence band, Taiwan 50 index.

I. INTRODUCTION

This dissertation can be divided into the two aspects of stock price prediction and Wavelet/HHT confidence bands. The AR(2) model of time series will be used for stock price prediction, and the primary condition for AR(2) prediction is the stationary state time series. However, in general the data of financial stock prices will not be presented in stationary state. Therefore, the two EMD methods in DWT and HHT of Daubechies

Wavelet Function will be used for data pre-treatment prior to the prediction such that the data can be in the form of stationary state time series for facilitating the future stock price prediction. Take the underlying stocks of Taiwan 50 Index as an example, the DWT(db3) will be used for 3~5 times of transforms to eliminate the high frequency coefficients such that the data can be shown in stationary

state. In terms of HHT, the EMD will be used to gradually obtain IMF based on repeated screening process, and the sum of the final two items of IMF will be regarded as the stock price trend. The data will also be shown in stationary state. The data after transform can then be used for stock price prediction based on AR(2). In the end, the K-S test and root mean square error will be used to examine whether or not the prediction residual is in normal distribution, and to conduct comparison between the accuracies of two transform methods, respectively. In terms of the Wavelet/HHT confidence bands, DWT and HHT methods will be used to obtain the upper and lower envelopes, based on which the data points in 10 transaction days will be predicted by AR(2) model. The data points predicted by such upper and lower envelopes will be connected by broken line, and the area surrounded by this broken line is called the Wavelet/HHT confidence band. In the end the statistical confidence bands and the size of their comparison zone will be used to compare the numbers of data points of true values of the two transform methods fallen within the Wavelet/HHT confidence band.

II. DWT AND HHT

The study in this dissertation is mainly about data pre-treatment and noise elimination based on third-order Daubechies Discrete Wavelet Transform. For the detailed Wavelet theories please refer to [1]-[3]. Hilbert−Huang Transform is also known as HHT [4]. In this dissertation the result of HHT has been used to establish the model for time series data of financial stock prices, based on which the further stock prices will be predicted.

A. Intrinsic Mode Function

Any given data meeting the following two conditions can be called an intrinsic mode function (IMF):

[1] The sum of numbers of local maximum values and local minimum values must be equaled to the number of values passing through zero point; otherwise the difference between these two numbers can only be 1.

[2] At any given time, the average value of upper envelope defined by local maximum value and the lower envelope defined by local minimum value must be zero.

B. Empirical Mode Decomposition

The empirical mode decomposition is for gradually obtaining IMF based on repeated screening process. Take the signal $x(t)$ as an example, the screening process is as shown below with a detailed graphical description:

Step 1: All local maximum values and local minimum

values in signal $x(t)$ should be obtained, and they should be connected by Cubic spline to form the upper envelope and lower envelope. Then the average value of upper and lower enveloped should be obtained as the average envelope $m_1(t)$ (see Fig. 2).

Step 2: The fist component can be obtained by subtracting the average envelope from the original signal as $h_1(t) = x(t) - m_1(t)$.

Step 3: $h_1(t)$ must be examined to see if it meets the condition of IMF. If not, return to Step 1 for a second screening with $h_1(t)$ as the original signal, such as

$$h_2(t) = h_1(t) - m_2(t)$$

after k times of repeated screenings,

$$h_k(t) = h_{k-1}(t) - m_k(t)$$

until $h_k(t)$ has met the conditions of IMF in order to obtain the first IMF component as $C_1(t)$.

Step 4: The first residual can be obtained by subtracting the first IMF component from the original signal as shown in the following equation:

$$R_1(t) = x(t) - C_1(t)$$

Step 5: Step 1 to Step 4 will be repeated with $R_1(t)$ as the new data in order to obtain the second residual $R_2(t)$. And it will be repeated for n times such as:

$$R_2(t) = R_1(t) - C_2(t)$$
$$R_3(t) = R_2(t) - C_3(t)$$
$$\vdots$$
$$R_n(t) = R_{n-1}(t) - C_n(t)$$

When the nth residual $R_n(t)$ has become monotonic function which cannot be decomposed into IMF, the process of EMD is completed. The original signal can be obtained by summing up all components of IMF and the last residual, thus explaining the applicability of HHT (see Fig. 3).

During the process of IMF calculation, there are two stopping criteria for Step 3:

1) When the number of peak values of signal and the number of values passing through the zero point are identical, it can be stopped.

2) $SD_k = \dfrac{\sum_{t=0}^{T} \left| h_{k-1}(t) - h_k(t) \right|^2}{\sum_{t=0}^{T} h_{k-1}^2(t)}$, if SD_k is less than the previous SD_{k-1}, and then the screening process can be stopped. Otherwise it shall continue. SD_k is actually the calculation of standard deviation, which can be regarded as the comparison of variation of fluctuation between functions. If the variation of SD_k is less than the previous one, the function will begin to converge, such that it is set to be the stopping criterion.
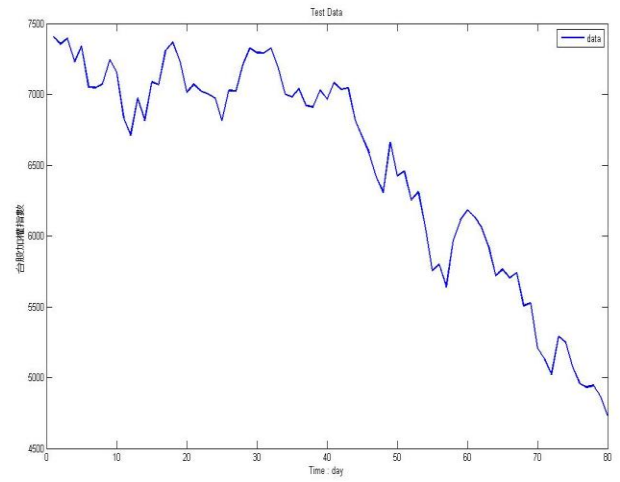

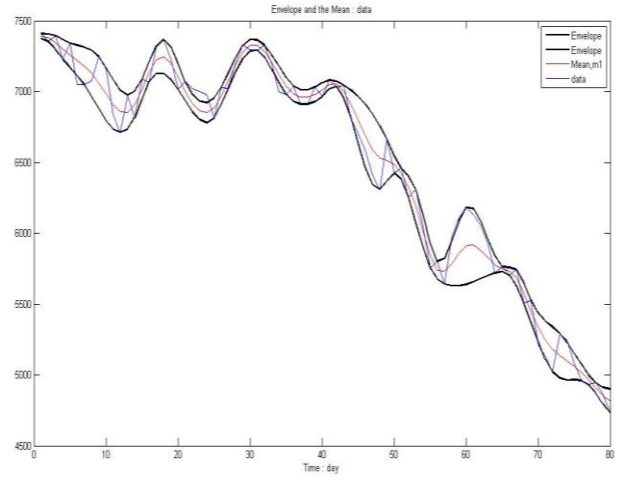Fig. 1. Taiwan capitalization weighted stock index (2001/03/23~2005/04/04).

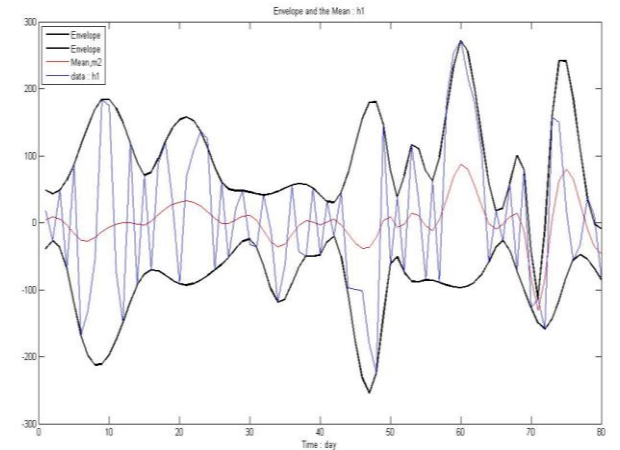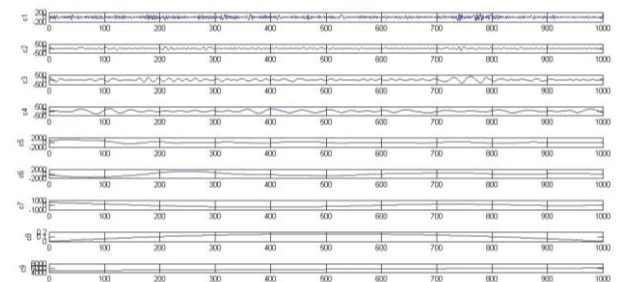Fig. 2. Average envelope $m_1(t)$.

Fig. 3. IMFs.

Fig. 4. IMFs.

The relationship between IMF component and original

signal will then be illustrated (Fig. 4-Fig. 6). It can be found that the original signal can be obtained by summing up all IMF components and the last residual as shown in the equation below:

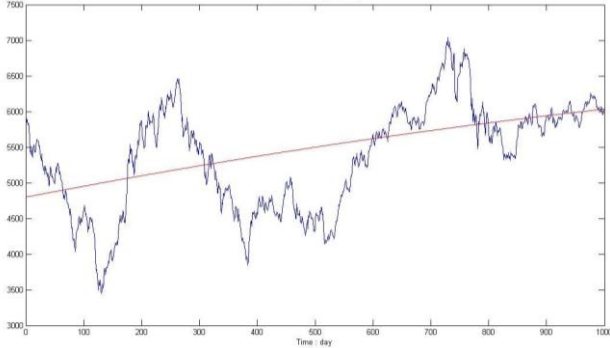$$x(t) = C_1(t) + C_2(t) + \cdots + C_n(t) + R_n(t)$$
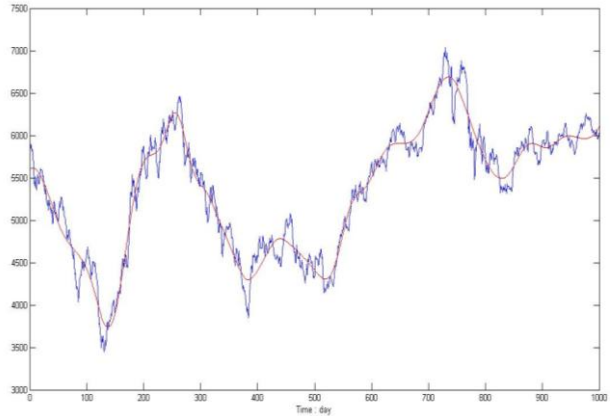

Fig. 5. 9th IMF.

Fig. 6. 5th IMF+…+9th IMF.

III. TIME SERIES

It has been pointed out in [5], [6] that the so-called time series data is the data recorded at different time points. Based on the data collection time period, it can be divided into the low frequency data (such as annual data, quarterly data, and monthly data) and high frequency data (such as weekly data, daily data, and individual transaction data within a day). Some time series appear to follow a fixed trend, while some don't. Serial correlation can be found in time series data, which means there is correlation between the data in this period with the data before or after this period. For example, the stock price today is bound to be under the influence of the price in the past, while it will definitely affect the price in the future. The moments of concern in random sample data are nothing more than mean and variance:

$$E(y_t) = \mu_t, \quad Var(y_t) = \sigma_t^2$$

However, with the serial related special structure of time series data, we are particularly interested about the connection among the past, present and future of data. Therefore, the autocorrelation function has been used to characterize the serial correlation of time series data.

In this study the experimental samples are samples of

concave curves, and DWT and HHT methods have been used for data pre-treatment such that the final trend is a concave curve. Therefore, AR(2) has been used as the prediction model in this study.

The second-order autoregressive model is to use the data in the first two periods of prediction as the explanatory variables, while the stationary state AR(2) model is as shown below:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t$$

$$\varepsilon_t \sim WN(0, \sigma^2), E(\varepsilon_t y_{t-1}) = 0, \forall j > 0$$

And the necessary and sufficient condition of this stationary state is:

$$\beta_1 + \beta_2 < 1, \quad \beta_2 - \beta_1 < 1, \quad |\beta_2| < 1$$

where β_1 and β_2 are called second-order autoregressive coefficients.

A. The Prediction by Second-Order Autoregressive Model

The second-order autoregressive model will be shown in matrix:

$$\begin{bmatrix} 1 & y_2 & y_1 \\ 1 & y_3 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & y_{t-1} & y_{t-2} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} y_3 \\ y_4 \\ \vdots \\ y_t \end{bmatrix}$$

By using the Left Division technique of Matlab:

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 & y_2 & y_1 \\ 1 & y_3 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & y_{t-1} & y_{t-2} \end{bmatrix} \backslash \begin{bmatrix} y_3 \\ y_4 \\ \vdots \\ y_t \end{bmatrix}$$

β_0, β_1, and β_2 can be obtained. And then the autoregressive coefficient can be used to predict the data in the next period:

$$\begin{bmatrix} 1 & y_t & y_{t-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = y_{t+1}$$

IV. EXPERIMENTAL METHODS AND RESULTS

Take the daily closing prices based on concave trend of 50 indicative companies in Taiwan 50 Index as the example, the time series of data can be shown as $y(n)$, $=1, 2, …, N$, and the pre-treatment of $y(n)$ data will be conducted by using the Empirical Mode Decomposition (EMD) of third-order Daubechies Wavelet and HHT Transforms. All 50 items of data will occupy too much space of this dissertation, so only the experimental results of stock price of Yuanta Financial Holding Co., Ltd. will be shown here. However, the comparison of all 50 items of data will be shown.

B. *Experimental Method and Design*

Method (1): The transform of *n-2* order will be applied to the wavelet basis function db3, where *n* is the number of IMF components of HHT. a_{n-2} is used as the main trend of this data point, based on which the second-order autoregressive model will be established in order to predict the data points in the next 10 transaction days by daily forecast method.

The most drastic fluctuation can be found in the first noise data (d_1) such that it can be regarded as the main noise data. A new set of data points will be obtained by subtracting the first noise data (d_1) from the original data, and then the upper and lower envelopes will be obtained, based on which the second-order autoregressive model will be used to predict the data points in 10 transaction days. The data points predicted by such upper and lower envelopes are connected by a broken line, and the area surrounded by this line can be regarded as the wavelet confidence band. It is hoped that the range of variation of future data points can be predicted, and the actual stock prices in the future can fall within this wavelet confidence band.

Method (2): the empirical mode decomposition method of HHT will be used to decompose data points into *n* sets of intrinsic mode function components imf(1), imf(2), …, and imf(*n*), and the sum of the final two sets of intrinsic mode functions shown as imf(*n*)+ imf(*n*-1) will be regarded as the main trend of data points, based on which the second-order autoregressive model will be applied to predict the data points of the next 10 transaction days by daily forecast method.

The most drastic fluctuation can be found in imf(1) among all IMF components such that it can be regarded as the main noise data. A new set of data points will be obtained by subtracting this imf(1) component from the original data, and then the upper and lower envelopes will be obtained, based on which the second-order autoregressive model will be used to predict the data points in 10 transaction days. The data points predicted by such upper and lower envelopes are connected by a broken line, and the area surrounded by this line can be regarded as the HHT confidence band. It is hoped that the range of variation of future data points can be predicted, and the actual stock prices in the future can fall within this wavelet confidence band.

C. *Experimental Results*

1) *The comparison between DWT and HHT*

With the 50 indicative companies in Taiwan 50 Index as the samples for our prediction, the results have indicated that DWT prediction performs better than HHT prediction in the first week with a share of 31/50, yet HHT prediction performs better than DWT prediction in the second week with a share of 29/50. As for the two-week prediction, the performances of DWT and HHT predictions are similar. As for the true values fallen within the Wavelet/HHT confidence bands established by DWT or HHT method in the first week, the second week, or during the two-week time span, during the first week there are 18 stocks with the same number data points fallen within both bands, and for 26 out of the rest 32 stocks the number of data points fallen

within the DWT band is greater; during the second week, there are 24 stocks with the same number data points fallen within both bands, and for 15 of out the rest 26 stocks the number of data points fallen within the DWT band is greater; and during the two-week time span there are 13 stocks with the same number data points fallen within both bands, and for 24 out of the rest 37 stocks the number of data points fallen within the DWT band is greater. We list details in Table I.

TABLE I: TWO-WEEK PREDICTION ON TAIWAN 50 INDEX

DWT and HHT	RMSE			Confidence band		
	First week	Second week	2 weeks	First week	Second week	2 weeks
Data points fallen within the DWT band is greater than HHT	31/50	21/50	25/50	26/32	15/26	24/37
Data points fallen within the HHT band is greater than DWT	19/50	29/50	25/50	6/32	11/26	13/37
The same				18	24	13

2) *The Stock Price of Yuanta Financial Holding Co., Ltd.*
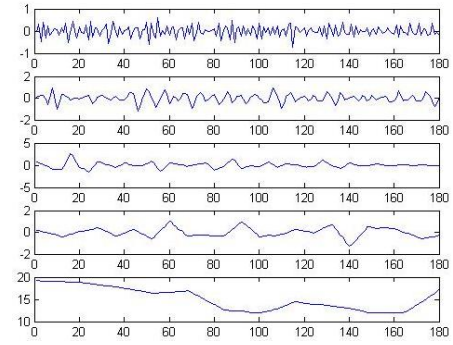

Fig. 7. DWT (db3, order 4) on Yuanta Financial Holdings Co.
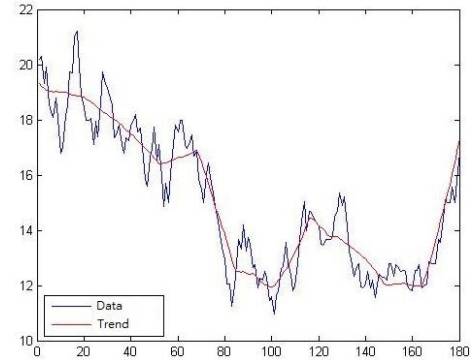

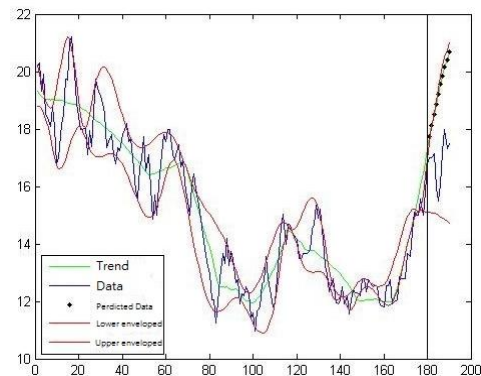Fig. 8. DWT (db3, order 4), a_4 as the trend of Yuanta Financial Holdings Co.

Fig. 9. DWT prediction on the stock price of Yuanta Financial Holding Co., and wavelet confidence bands.

This data has 6 sets of IMF components in accordance with the EMD method of HHT. Thus DWT will be applied for fourth-order transform of wavelet basis function db3 with a_4 as the trend of this data (Fig. 7). DWT decomposition will be used for the pre-treatment of noise elimination, and the second-order autoregressive model will be applied to a_4 (Fig. 8). After the Left Division technique of Matlab:

$$\beta_0 = 0.0409 \quad \beta_1 = 1.9680 \quad \beta_2 = -0.9706$$

where the second-order autoregressive coefficients are:

$$\beta_1 + \beta_2 < 1, \quad \beta_2 - \beta_1 < 1, \quad |\beta_2| < 1$$

such that it can be regarded as a stationary state time series.

The Fig. 9 shows that DWT prediction on the stock price of Yuanta Financial Holding Co., and behavior of wavelet confidence bands.

The stock price of Yuanta Financial Holding Co., Ltd. can be decomposed into 6 sets if IMF by EMD method of HHT, and the sum of the final two sets of IMFs can be regarded as the trend of this data, to which the second-order autoregressive model will be applied (Fig. 10-Fig. 11). After the Left Division technique of Matlab:

$$\beta_0 = 0.0218 \quad \beta_1 = 1.9973 \quad \beta_2 = -0.9987$$

where the second-order autoregressive coefficients are:

$$\beta_1 + \beta_2 < 1, \quad \beta_2 - \beta_1 < 1, \quad |\beta_2| < 1$$

such that it can be regarded as a stationary state time series.
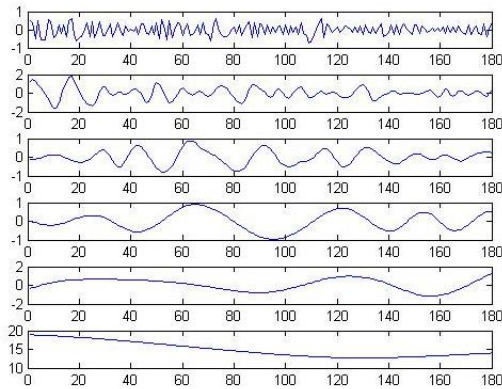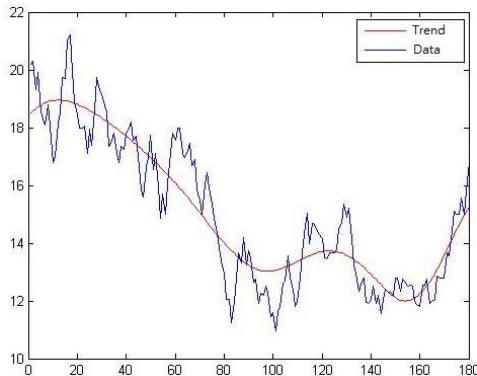

Fig. 10. HHT on Yuanta Financial Holdings Co.

Fig. 11. Trend (HHT), and the stock price of Yuanta Financial Holdings Co.

The Fig. 12 shows that HHT prediction on the stock price of Yuanta Financial Holding Co., and behavior of HHT confidence bands.
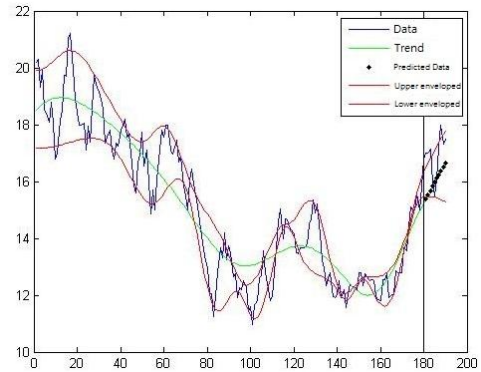

Fig. 12. HHT prediction on the stock price of Yuanta Financial Holding Co., and wavelet confidence bands.

D. Residual Test-the Stock Price of Yuanta Financial Holding Co., Ltd.

The K-S test will be used to examine whether or not the residuals are in normal distribution. The statistics software MINTAB and SPSS will be used to compose the regression line distribution and K-S test table for residual estimation. The residual distribution and K-S test statistics will be used to verify whether or not the residuals are in normal distribution.

1) The predicted residual of stock price of Yuanta Financial Holding Co., Ltd. after DWT pre-treatment will be examined (Fig. 13).
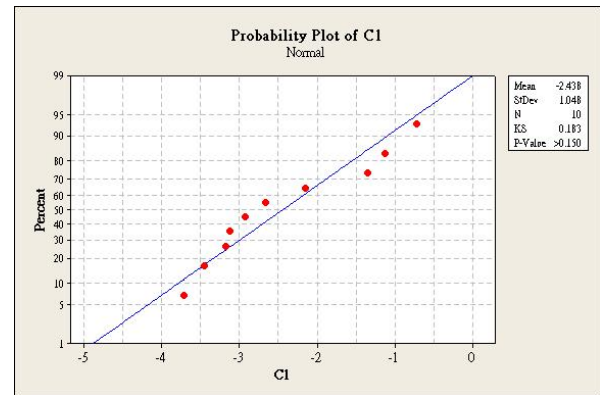

Fig. 13. residual of stock price of Yuanta Financial Holding Co., Ltd. (after DWT pre-treatment).

Based on the statistics of K-S test

$$d_{10} = 0.183 < D_{10, \, \alpha=0.05} = 0.41$$

Thus it is believed that the residuals between the original signal and the signal after DWT treatment are in normal distribution, and the mean and variance of residuals can be obtained.

2) The predicted residual of stock price of Yuanta Financial Holding Co., Ltd. after HHT pre-treatment will be examined (Fig. 14.).

Based on the statistics of K-S test

$$d_{10} = 0.200 < D_{10, \, \alpha=0.05} = 0.41$$

Thus it is believed that the residuals between the original signal and the signal after HHT treatment are in normal

signal and the signal after HHT treatment are in normal distribution, and the mean and variance of residuals can be obtained.
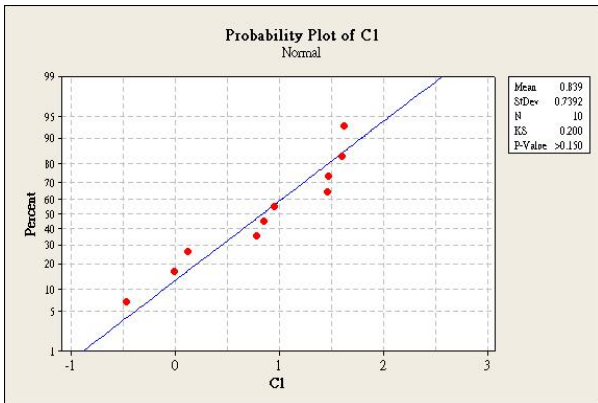

Fig. 14. Residual of stock price of Yuanta Financial Holding Co., Ltd. (after HHT pre-treatment).

Based on the statistics of K-S test

$$d_{10} = 0.205 < D_{10,\ \alpha=0.05} = 0.41$$

Thus it is believed that the residuals between the original signal and the signal after DWT treatment are in normal distribution, and the mean and variance of residuals can be obtained.

REFERENCES

[1] W.-C. Shan, *First Concepts of Wavelets*, Chuan Hwa Book Co., Ltd. 1999.
[2] A. Boggess and F. J. Narcowich, *A first course in wavelet with fourier analysil*, Prentice Hall, 2001.
[3] S. Mallat, "A theory for multi-resolution approximation: the wavelet approximation," *IEEE Trans.*, vol. 11, 1989, pp. 674-693.
[4] N. Huang, *A plea for Adaptive data analysis*, 1998.
[5] X.-S. Chen, *The application of time series analysis for macro-economy and finance*, Prentice Hall, 2007.
[6] L. M. Liu, *Time series analysis and forecasting*, Scientific Computing Associates, 2006.

H. R. Ke received his B.S., M.S. degrees in mathematics from National Cheng Kung University, Feng Chia University in 1997 and 1999, respectively. Since 1999, he has been a teacher in the Beidou Junior High School, Changhua, Taiwan. His research interests include wavelet analysis.

K. C. Wang received his B.S., M.S. degrees in mathematics from National Cheng Kung University, Feng Chia University in 2000 and 2003, respectively. He is a researcher in Ph.D. Program in Mechanical and Aeronautical Engineering. Since 2010, he has been a lecturer in the Department of Applied Mathematics at Feng Chia University, Taiwan. His research interests include frame theory and wavelet analysis.

C. I Yang received his B.S., M.S. degrees in mathematics from National National Taichung University of Education, Feng Chia University in 1999 and 2011, respectively. Since 1999, she has been a teacher in the Dongxing elementary school, Taichung, Taiwan. Her research interests include Neural Networks and wavelet analysis.

K. F. Chang received his B.S., M.S. and Ph.D. degrees in mathematics from National Taiwan Normal University, National Tsing Hua University and the University of Texas at Austin in 1979, 1981 and 1993, respectively. Since 1998, he has been a professor in the Department of Applied Mathematics at Feng Chia University, Taiwan. His research interests include approximation theory and wavelet analysis.