Journal of Theoretical and Applied Information Technology
10th March 2013. Vol. 49 No.1

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

PRICE TIME SERIES LONG MEMORY ANALYSIS AND PREDICTION STUDY

[1]XIAOBING CHENG, [2]MINGPING XIA, [3]RUIKAI PAN, [4]DAOPING WANG

[1]Ph.D., College of Automation, Beijing Union University, Beijing, 100101

[2]Lecture, College of Automation, Beijing Union University, Beijing, 100101

[3]Lecture, School of Chemistry and Biology Taiyuan University of Science and Technology, Taiyuan, 030021

[4]Prof., School of Economic and Management, University of Science & Technology Beijing, 10083

E-mail: [1]xiaobingcheng@sina.com, [2]zdhtmingping@buu.edu.cn, [3]dpwang@ustb.edu.cn

ABSTRACT

The R/S test has been extensively used in testing the long memory of financial time series, but little attentions have been paid on its validity. The paper sets the chemical raw materials styrene price time series as an example, to test the stable of the price series. It indicates that we should give prudent explanation for the R/S test, and then establish the ARFIMA model to determine the data generation process with fractal characteristics. Thus the fractal theory can be used to describe the price time series and provide theoretical support for the price time series forecasting.

Keywords: *Validity, Long memory, Forecasting*

1. INTRODUCTION

Natural phenomenon and social economic phenomenon are both regular dialectical development courses. All movement has certain inertia, and it shows a kind of dynamic of the system, namely memory [1]. After Hurst found long memory of hydrological time series from the tidal data, the research of long memory has been caused widely public concern such as fluid science, meteorology and geophysics and so on. Economists found that we can't ignore the correlation among the distant time interval measurement of a large number of economic time series, such as stock price, economic growth rate, inflation rate, oil price and GDP figures etc, the time series showed a characteristic of "long memory faculty" [2].

Long memory shows that time series exists a continuous long-term dependency among the distant time interval measurement [3]. When the delay order number k was larger, time series has a correlation in the time value and t-k time value, and this link is often measured by the autocorrelation coefficient of the series, and the memory extend of the series can be judged by attenuation way of the autocorrelation coefficient curve. For a stationary time series, if it is a financial time series, it will mean the failure of random walk model and efficient market hypothesis and many other classical financial theories [4]. If the dynamic dependent structure shows a long memory or long-term dependent relationship, the history information of the series will help to predict the future change.

2. LONG MEMORY TIME SERIES

2.1 Long memory analysis

Definition 1: we assume time series $\{X_t\}$ has self correlation function ρ_τ, and τ is the lagging number. If ρ_τ satisfies the condition: $\lim_{T\to\infty}\sum_{\tau=-n}^{n}|\rho_\tau|\to\infty$ and then $\{X_t\}$ is called long memory time series (Mcleaod and Hipel).

The autocorrelation function of long memory process is not decaying at an exponential rate, but with hyperbolic velocity decaying slowly, so that its autocorrelation function can't be added.

Inspecting the long memory, the statistical methods we usually use include the correlation coefficient method, the classical R/S test (or heavy rescaled range test), the modified R/S test (MR/S), KPSS method, logarithmic diagram method (GPH), and Gauss semi-parametric estimation method (GSP) etc, and the most extensive analysis method

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

is the classical method R/S. But R/S method usually tends to reject the original hypothesis of "the long memory does not exist" and it is in connection with principle of R/S test statistic [5].

The specific thought of R/S analysis method is: time series $\{X_t\}$ of the sample length T is divided into k son intervals of length $n(n \times k = T)$, and the average of n series observed values is $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ The range of each subinterval is defined $R(n)$, and the standard deviation $S(n)$. Take the statistical R/S [6]:

$$Q_n = R(n)/S(n) \quad (1)$$

The range $R(n)$ and the standard deviation $S(n)$ are respectively:

$$R(n) = \max_{1\le k\le n}\sum_{j=1}^{k}(x_j - \bar{x}_n) - \min_{1\le k\le n}\sum_{j=1}^{k}(x_j - \bar{x}_n),$$

$$S(n) = \left[\frac{1}{n}\max_{1\le k\le n}\sum_{j=1}^{k}(x_j - \bar{x}_n)^2\right]^{\frac{1}{2}}.$$

We can prove that $P\lim_{n\to\infty}(n^{-H}Q_n) = C$, C is a constant, and H is Hurst index, so we can get approximate estimate of H, $H = \ln Q_n / \ln n$.

In general, the R/S analysis method is described as follows (2).

$$(R/S)_n = C \cdot n^H \quad (2)$$

In type (2), R is rescaled range, S is the standard deviation, H is Hurst index, C is a constant, and n is sample observation number. On the (2) type of logarithmic, we get:

$$\log(R/S)_n = \log(C) + H\log(n) \quad (3)$$

Hurst index is in [0, 1] value, according to its value the time series can be divided into three different types [7]:

$H = 0.5$, indicating the correlation coefficient between past and future increment of series is zero, and namely it does not affect the future, and the incremental process is an independent random process, the series is the random walk and it is gradual process. The actual financial time series often has some characteristics like short data, big noise. The limited sample often makes value of random series H deviate from 0.5, so it needs to test the significance of the H index.

$0 < H < 0.5$, the system of this type is defined as the reverse persistent series, and also often referred to as "mean reversion ". H is relatively rare in this range, and then the autocorrelation coefficient is between 0.5 and 1. If the series was growth, then declining possibility in the next moment will be bigger; On the contrary when the past is decreasing, then increasing possibility in the next time will be bigger. Counter lasting effect depends on if H close to 0. When H is closer to 0, and this process has more frequent reversible, so it is more severe than fluctuation of Brownian motion.

$0.5 < H < 1$, the system of this type is the persistent or trend enhanced series. Then the autocorrelation coefficient is between 0 and 1. This indicates that if the past has the trend of growth, it means that this trend will continue in the future; conversely, the decreasing trend of the past means that the future continues to decrease. H is closer to 1, and the trend is more obvious; H is closer to 0.5, the trend is gradually becoming random. This kind of long-term memory makes random process present certain trend, and there is a certain positive correlation among the increment.

2.2 Fcaoresting model of long memory

Mandelbrot put forward the concept of "Fractional Vibrant Motion" in 1968, which laid the foundation for the long memory model. In fractal dimension difference noise FDN model, he ignored the short memory of the time series, and in order to make up for the deficiency of FDN model [8]. In 1981, Granger and Hosking combined the traditional auto regression moving flat Model ARMA (Auto Regressive Moving are Average, ARMA) and FDN model, created the autoregressive integrated moving average model ARFIMA (Auto Regressive Fractional Integrated Moving Average, ARFIMA), for the analysis of long-term relationship of time series. The main difference between FDN and ARFIMA model is that the residual series of FND is white noise, and residual series of ARFIMA model is a wider stationary series [9]. In this sense, FDN model is a special case of ARFIMA model. When $p = q = 0$, $ARFIMA(0, d, 0)$ model is the fractional difference noise model FDN, which can be used to describe long related characteristics of the sample data, and when $d = 0$, $ARFIMA(p, 0, q)$ is the process of $ARMA(p, q)$, which can be used to describe short range correlation of the sample data.

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

Assume stationary time series $\{x_t\}$ meets the difference equation [10]:

$$\phi(B)(1-B)^d(X_t-\mu)=\theta(B)\varepsilon_t \qquad (4)$$

In the equation B is the lag operator, $|d|<0.5$, $\{\varepsilon_t\}$ is white noise series, and μ is the mean of $\{x_t\}$. $\phi(B)$ is p-order stable autoregressive lag polynomial, $\theta(B)$ is q-order reversible mobile lag polynomial, all of its latent roots are outside the unit circle, and then $\{x_t\}$ meets the autoregressive integrated moving average model, recorded as $ARIMA(p,d,q)$, the number $p+q$ is used to describe the process of short memory characteristic, parameter d to represent the process of long memory characteristics [11].

The autocorrelation function of process of $ARFIMA(p,d,q)$ presents attenuation of hyperbolic rate, it can analyze and describe short-term memory and long-term memory of time series, parameter d reflects the long memory among the time series observed values, parameter p and q reflect the short memory among the time series observed values [12]. $ARFIMA(p,d,q)$ model is one of the most effective tools to analyze long memory time series, and ARFIMA model is more complex than the traditional ARMA model, and it is difficult when the data have a long memory and short memory at the same time.

The basic step of building $ARFIMA(p,d,q)$ model is:

Data preprocessing: the trends and fluctuations of the original series are removed.

Removing short memory factors: short memory factors is filtered to extrude long memory by building AR model.

Analysis long memory factors of the time series to achieve fractional differential: parameter d is estimated preliminarily by using polymerization variance method, regression residuals method [13] and cycle diagram method etc in this step [14]. Through the score difference we get series $\{w_t\}$ is:

$$w_t=(1-B)^d(x_t-\mu) \qquad (5)$$

$\Phi(B)w_t=\theta(B)a_t$, $\{w_t\}$ is zero mean $ARMA(p,q)$ series.

To fix $ARFIMA(p,d,q)$ p, q order: this is fixed order questions of the conventional $ARMA(p,q)$ model.

3. LONG MEMORY TIME SERIES

3.1 Long memory analysis

Styrene monomer belongs to relatively larger petrochemical products, and demand is very strong. It is produced generally by large petrochemical enterprise in China [15], its production scale is larger, and its downstream resin, rubber industry characteristics is a single small scale of enterprises, and the enterprise quantity is numerous. Coupled with styrene monomer in short supply, pricing is mastered by the seller. So it is very important to analyze and predict the change trend of styrene market price [16].

This paper chose 178 weeks on average market price (dollars) of styrene monomer from 2006 May to 2009 November as the research and analysis of sample, this period of styrene products price sequence is used as the sample data to analyze long memory property. The change trend is shown in Figure 1.

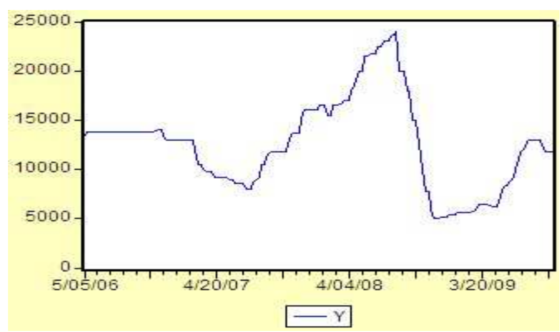


Figure1: May 2006-September 2009 Styrene Price Chart

Assume time series $\{x_t\}, t=1,2,\cdots,T$, and the mean of n observation values is $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=12579.216$, then we can get:

$$R(n)=\max_{1\le k\le n}\sum_{j=1}^{k}(x_j-\bar{x})-\min_{1\le k\le n}\sum_{j=1}^{k}(x_j-\bar{x})$$
$$=223931.691\text{-}(-52680.051)$$
$$=276611.742$$

$$S(n)=\sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_j-\bar{x})^2}=4625.712$$

$$H=\ln Q_n/\ln n=0.789$$

The approximate estimate of H is 0.789, and it is between 0.5 and 1. We can judge the time series

Journal of Theoretical and Applied Information Technology
10th March 2013. Vol. 49 No.1

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

sample data having the characteristic of stronger long memory, and it shows a strong correlation among the distant time series.

To test the stability of the original data, we get great correlation coefficient, and the coefficient of lag 16 period is still 0.413, existing obvious correlation. We differ the series $\{x_t\}$ and test the series ADF, the results is shown in figure 2.

ADF Test Statistic	-2.355090	1% Critical Value*	-3.4693
		5% Critical Value	-2.8782
		10% Critical Value	-2.5756

*MacKinnon critical values for rejection of hypothesis of a unit root.

Figure2: The ADF Test Results Of Time Series

From figure 2 we can see that the ADF test statistic is -2.35509, and it is respectively greater than the 3 different inspection levels which are 1%, 5%, 10%. Therefore, the original series is not stationary, statistical diagram of the original series $\{y_t\}$ is shown in figure 3.

The mean of value styrene is 12579.22 in the sample period, and the standard deviation is 438.76, the skewness is 0.397801, the left peak is 2.793022. As we can see from Figure 3, styrene price has a peak characteristic. The associated probability p is 0.081579 of J-B test, indicates the confidence level of at least 95% to reject the null hypothesis, and namely the sequence is not subject to the normal distribution.

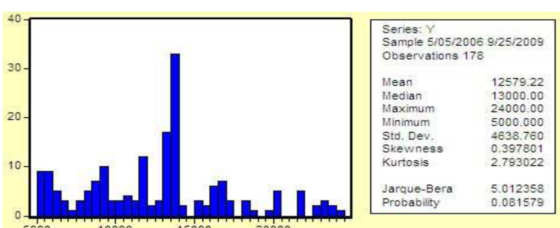


Figure3: Diagram Of The Original Series $\{y_t\}$

We can find that styrene price fluctuates with segmentation and partition from Figure 1. Two points respectively have a relatively acuteness wave motion in February of 2007 to 2009, and we also found that some period of time are continuous high, some period of time are continuous low. We take first-order difference to original series of $\{y_t\}$, and get the difference series chart shown in figure 4.

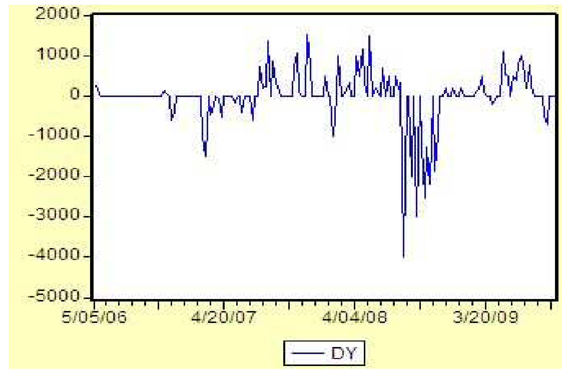


Figure4: First Order Difference Series

From Figure 5, the sample period styrene mean is closer to -9.60, the standard deviation is 654.0563, the skewness is -2.444, the left peak is 14.3463. As we can see from Figure 3, styrene price has a peak characteristic. The associated probability p is 0.081579 of J-B test, indicates the confidence level of at least 99% to reject the null hypothesis, and namely and namely the series after difference obeys normal distribution. The autocorrelation coefficient of difference series reduces greatly. So differ the difference series $\{dy_t\}$ and do ADF test for the series, test results as shown in figure 6.

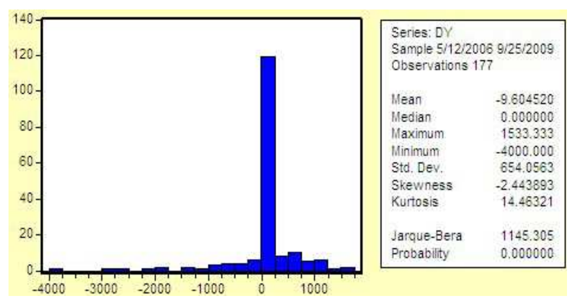


Figure5: First Order Difference Series Chart

ADF Test Statistic	-2.438973	1% Critical Value*	-3.4695
		5% Critical Value	-2.8783
		10% Critical Value	-2.5756

*MacKinnon critical values for rejection of hypothesis of a unit root.

Figure6: First Order Difference Series ADF Test

From figure 6 we can see that the ADF test statistic is -2.438973, and it is respectively greater than the 3 different inspection levels that are 1%, 5%, 10%, indicating at least 99% confidence level to reject the null hypothesis, the series dy_t existing unit root, namely not smooth series. Then we give the series two order difference, get two order difference series $dy2$. The trend diagram of two order difference series is shown in figure 7.

Journal of Theoretical and Applied Information Technology
10th March 2013. Vol. 49 No.1

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

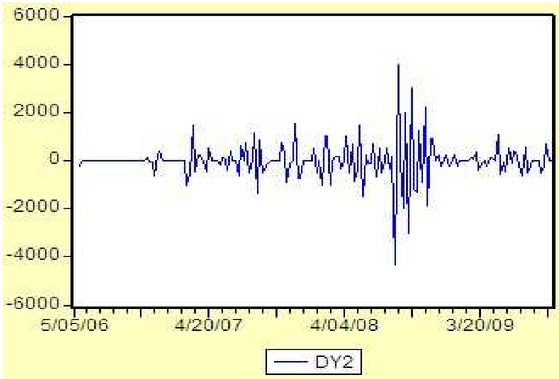


Figure7: Two Order Difference Series

We can find that styrene price fluctuates with segmentation and partition from Figure 1.Two points respectively have a relatively acuteness wave motion in February of 2007 to 2009, and it is also found that some period of time are continuous high. We get the autocorrelation coefficient and partial correlation coefficient from the difference series, two order difference series chart as shown in figure 8.

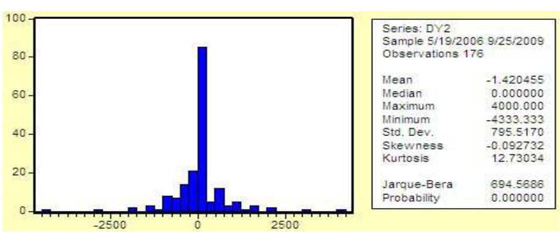


Figure8: Two Order Difference Series Chart

From Figure 8, we can know that styrene mean is closer to -1.42 in sample period, the standard deviation is 795.517, the skewness is -0.092732, the left peak is 12.73034, the styrene price has a peak characteristic. The associated probability p is 0 of J-B test, indicates at least 99% confidence level to reject the null hypothesis, and namely the series after difference obeys normal distribution. The autocorrelation coefficient of difference series reduces greatly. So difference to the difference series $\{dy2_t\}$ and ADF test for the series, the test results is shown in figure 9.

ADF Test Statistic	-7.083835	1% Critical Value*	-3.4697
		5% Critical Value	-2.8784
		10% Critical Value	-2.5757

*MacKinnon critical values for rejection of hypothesis of a unit root.

Figure9: Two Order Difference Series ADF Test

Figure 9 we can see that the ADF test statistic is -7.083835, it is respectively smaller than the 3 different inspection levels that are 1%, 5%, 10%, indicating that $dy2$ rejects the null hypothesis at least 99% confidence level, the series $dy2_t$ doesn't exists unit root, namely smooth series. And $d = 2$.

3.2 Model establishment and verification

For $ARFIMA(p,d,q)$ model, we can use the sample autocorrelation function and the sample partial autocorrelation function of truncated to determine the order of the model. If the stationary time series of partial correlation function is censored, the autocorrelation function is heavy-tailed, we can conclude that this sequence adapts to $AR(p)$ model; If partial correlation function of the stationary time series is trailing and the autocorrelation function is truncated, we can conclude that this series adapts to $MA(q)$ model; If partial correlation function and the autocorrelation function of the stationary time series are tailing, this series adapts to $ARFIMA(p,d,q)$ model. Graphs of autocorrelation and partial autocorrelation have characteristics of tailing and sinusoidal tends to zero. According to the Box-Jenkins model identification method, we use $ARFIMA(p,d,q)$ model to fit. After $k=5$ partial correlation coefficient quickly approaches to 0, so $p=5$; Autocorrelation coefficient wasn't obviously to be 0 when $k=1$, and it also shows significant differences with 0 when $k=1$, and when $k=2$ it soon tends to 0, and we can consider $q=1$. With the help of Eviews, we can get the $ARFIMA(5,2,1)$, the results shown in figure 10.

Dependent Variable: DY2
Method: Least Squares
Date: 04/22/10 Time: 18:14
Sample(adjusted): 6/23/2006 9/25/2009
Included observations: 171 after adjusting endpoints
Convergence achieved after 6 iterations
Backcast: 6/16/2006

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-1.296475	0.210636	-6.155056	0.0000
AR(2)	-1.004623	0.214074	-4.692881	0.0000
AR(3)	-0.627823	0.198506	-3.162739	0.0019
AR(4)	-0.412112	0.166282	-2.478395	0.0142
AR(5)	0.035763	0.130886	0.273241	0.7850
MA(1)	0.553609	0.198217	2.792943	0.0058

R-squared	0.553269	Mean dependent var	-5.32E-15
Adjusted R-squared	0.539731	S.D. dependent var	806.9774
S.E. of regression	547.4785	Akaike info criterion	15.48298
Sum squared resid	49455905	Schwarz criterion	15.59321
Log likelihood	-1317.795	F-statistic	40.86990
Durbin-Watson stat	2.000418	Prob(F-statistic)	0.000000

Inverted AR Roots	.09+.74i	.09 -.74i	.08	-.78 -.48i
	-.78+.48i			
Inverted MA Roots	-.55			

Figure10: ARIMA(P, D, Q) Initial Correlation Coefficient Calculated Value

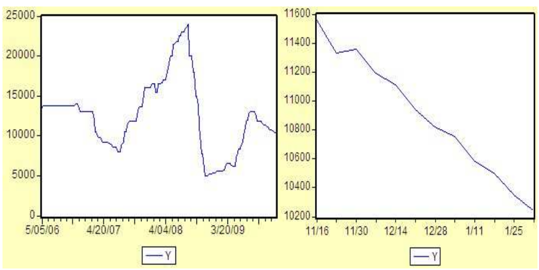
According to the fitting results, we eliminate AR (5), and the values of other p are remarkable at the 5% level. Then we get the forecasting model as follows:

$$\begin{aligned} dy2_t = & -1.296475dy2_{t-1} - 1.004623dy2_{t-2} \\ & -0.627823dy2_{t-3} - 0.412112dy2_{t-4} \\ & +d\varepsilon_t + 0.553609d\varepsilon_{t-1} \end{aligned} \quad (6)$$

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

On the sequence of Q-statistic test, all the Q values are smaller than the test level of 0.05 χ^2 distribution critical value. We can draw a conclusion: random error series of the model is a white noise series, and model test is passed.

The trends of two order difference series back into the actual series is shown in figure 11.



(A) (B)

(A) The Overall Trend Diagram

(B) Trend Chart Of The Next 12 Weeks Period

Figure11: Styrene Price Series Trend Chart

4. CONCLUSION

The test methods of time series long memory: empirical discriminates method based on the sample autocorrelation, R/S test, MR/S test, GYH test etc. R/S test is more sensitive on short memory of time series, and it tends to reject the original hypothesis of the long memory which doesn't exist. Although MR/S test can improve sensitivity and robustness of the model for short memory, but the effect depends on suitable selection of the truncation parameter.

Because a large number of time series characteristics is different from I(0) series and I(1) series, so we can't use traditional methods to study it. Mandelbrot (1968) put forward the concept of "Fractional Brownian Motion" which laid the foundation for long memory modeling. The autocorrelation function of process $AFIMA(p,d,q)$ presents attenuation of hyperbolic rate, and at the same time we can describe and analyze the short-term memory and long-term memory. Model parameter d reflects the long memory among the time series observed values, and model parameter p, q reflect short memory among the time series observed values.

As the example of the important chemical raw materials styrene market price, we study on the long memory of the price series. The research is:
as short memory time series is concerned, R/S test tends to accept the existence of long memory, and we should cautiously analysis results of the R/S test. residual error series of the ARFIMA model is a more widely stationary series. In this sense, the FDN model is a special case of the ARFIMA model. That is when $p=q=0$, $ARFIMA(0,d,0)$ model is the fractional difference noise model FDN, and it can be used to describe the long correlation characteristics of the sample data. When $d=0$, $ARFIMA(p,0,q)$ is the process of $ARMA(p,q)$, and it can be used to describe the short range correlation of the sample data.

5. ACKNOWLEDGEMENT

Supported By National Natural Science Foundation of China (71172169/G0212).

REFRENCES:

[1] Hurst, H. E. "Long-term storage capacity of reservoirs", *Transactions of the American Society of Civil Engineers*, Vol. 35, No. 16, 1951, pp. 700-779.

[2] YouChuan Xu, "The effectiveness of long memory of Financial Time series R_S test", *China Management Science*, Vol. 11, No. 18, 2010, pp. 204-208.

[3] JunGang Wang, Yong Wang, and ZhiPu Huo, "Analysis on Effect of Excitation Frequency on Nonlinear Characteristics of a Gear Pair System Based on numerical Simulation", *Journal of Theoretical and Applied Information Technology*, Vol. 47, No. 1, 2013, pp. 362-368.

[4] Campbell, John Y, and John H Cochrane, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior", *Journal of Political Economy*, Vol. 10, No. 7, 2009, pp. 205-251.

[5] R.S.Aarthi, D. Muralidharan, P. Swaminathan, "Double Compression of Test Data Using Huffman Code", *Journal of Theoretical and Applied Information Technology*, Vol. 39, No. 2, 2012, pp. 104-113.

[6] Granger, C.W.J, Joyeux, R, "An introduction to long memory time series models and fractional differencing", *Journal of Time Series Analysis*, Vol. 31, No. 1, 1980, pp. 15-39.

[7] WenHui Xia, JuanHong Liu, Hong Zeng, "Application of Nested Cost Optimization Models to TSPD", *Journal of Theoretical and Applied Information Technology*, Vol. 47. No. 1, 2013, pp. 251-255.

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

[8] Taqqu M S, Teverovsky V, and Willinger W, "Estimations for long-range dependence: an empirical study", *Fractals*, Vol. 3, No. 4, 2008, pp. 765-788.

[9] Peng K K. "Mosaic organization of FND model", *Physical Review E*, Vol. 49, No. 56, 2011, pp. 1685-1689.

[10] Safwan AL Salaimeh, "A New Model For Information Logistics System Architecture", *Journal of Theoretical and Applied Information Technology*, Vol. 28. No. 1, 2011.

[11] R. Jayabrabu, Dr. V. Saravanan, and Prof. K. Vivekanandan, "Software Agents Paradigm in Automated Data Mining for Better Visualization Using Intelligent Agents", *Journal of Theoretical and Applied Information Technology*, Vol. 39, No. 2, 2012, pp. 167-177.

[12] Liliana and Togar Alam Napitupulu, "Artificial Neural Network Application in Gross Domestic Product Forecasting an Indonesia Case", *Journal of Theoretical and Applied Information Technology*, Vol. 45, No. 2, 2012, pp. 410-415.

[13] Yousef Khaleel and Riza Sulaiman "A System Development Methodology for ERP System in SMES of Malaysian Manufacturing Sectors", *Journal of Theoretical and Applied Information Technology*, Vol. 47, No. 2, 2013, pp. 504-513.

[14] G. PREETHI and B. SANTHI, "Stock Market Forecasting Techniques: A Survey", *Journal of Theoretical and Applied Information Technology*, Vol. 46, No. 1, 2012, pp. 24-30.

[15] Yang Li, Chen Lixue and Gen Xinyu, "Application of Factor Neural Network in Multi-expert System for Oil-gas Reservoir Protection", *Journal of Theoretical and Applied Information Technology*, Vol. 46. No. 1 – 2012, pp 303 - 308

[16] Bin Wang and Jing Wang, "On Price and Service Competition with Substitutable Goods", *Journal of Theoretical and Applied Information Technology*, Vol. 46, No. 1, 2012, pp. 355-364.