ISSN: 2688-9323

Finance Forecasting in Fractal Market Hypothesis

Wangke Lin

Zhejiang College of Security Technology, Zhejiang, China

Abstract

Based on Fractal Market Hypothesis and the chaotic dynamics theory, the Swedish capital markets have been tested by Hurst Exponent and the largest Lyapunov Exponent which show the Swedish markets have fractal and chaotic behavior. It is helpful to find a better model to forecasting these market, ARIMA, BP neural network and the hybrid mode which combined ARIMA and BP (Backpropagating) neural network model, are basic models to forecasting these markets. The results show the hybrid model can be an effective way to improve forecasting accuracy than the ARIMA and BP neural network model.

Keywords

Hurst Exponent; Lyapunov Exponent; ARIMA; BP Neural Network.

1. Introduction

The world is not orderly, nature is not orderly, and the capital markets are also not orderly. They contain fractal characteristics, chaotic behaviors and non-linear, which make them extremely complex and unpredictable. There are some of the most challenging problems in Capital markets are chaotic behaviors and non-linear, complication and uncertainty, unexpected booms and crashes. A number of people are trying to explain these problems and understand how the capital flows and why. The E. Peters, 1996 proposed Fractal Market Hypothesis (FMH). Theorizes that "(1) in the different time horizons, all investors are being included in the market stability and liquidity. (2) though the changing of the market information, it takes time for investors to reflect or just wait for more information, (3) there are some information maybe can't reflect the market price, (4) market price shows the long-term memory and also indicates the long-term economic trends" (E. Peters, 1996). In the hypothesis of FMH, the market is non-Gaussian distribution and non-stationary process which means the information have historical correlation exist, in other words, finance have a long memory. Fractal Market Hypothesis is different to Efficient Market Hypothesis. The Efficient Market Hypothesis state that: "(1) All available information should be fully reflected in the current price of an asset. (2) The new information is combined in price independent and random. (3) All investors immediately react to new information. All these three conditions means that the future is unrelated to the past or the present." (E. Peters, 1996). While, most people do nothing before the trend is clearly showed and the information is been surely confirmed, but some people maybe react as soon as the information received. Because of most people confirming information in different time and not react it as soon as possible, which caused a biased random walk. Biased random walk were been called fractional Brownian motions by Mandelbrot (Mandelbrot 1968). Now they are called fractal time series. "While Fractal Market Hypothesis based on fractal and chaos theory. Chaotic and fractal system is a unconventional collage consists of deterministic and random process." (E. Peters, 1996).

There are a number of important characteristics indicate that, if the capital markets are nonlinear dynamic systems, according to the Fractal Market Hypothesis, the capital markets exist nonlinear dynamic systems, then we should expect: " first they are feedback system, what happened yesterday influences what happens today, in the other word there exist long-term correlations and trends. Second erratic (critical levels) markets under certain condition at certain times. Third the system is fractal, a time series of return that, at smaller increments of time will still look the same and will have similar statistical characteristics. Finally, there is sensitive dependence on initial conditions and less reliable forecast, it is extremely chaotic and minor events can cause major perturbation in final outcome." (E. Peters, 1996). However, due to the capital market's unpredictable movement, there is always existing risk to investment in capital market. Thus it would be highly valued and helpful to find an appropriate prediction model for capital market forecasting.

In this article, first Swedish capital markets have been selected to test if the Swedish capital markets have fractal and chaotic characteristics. Second, the Auto-regressive integrated moving average (ARIMA) model, BP neural network and the hybrid model have been chosen to predict, compared with the results of those three models the "true" or "best" model should be found out to the Swedish capital markets. A new strategy is to use C-C method (Kim 1999), to reconstruct the BP neural network and the hybrid model.

2. Method

2.1. Fractal and Fractional Dimension

2.1.1. Fractal

What is fractal? A fractal is an object in which the parts are in some way related to the whole. Fractals are self-referential, in another word, they have a large degree of similarity within themselves. Look at trees, they are perceived nature fractals. Trees have similar branches. Fractal geometry was first mentioned by Benoit Mandelbrot (1975). The self-similar quality is the defining characteristic of fractals. Fractal geometry just right can describe natural shapes, far more "pure" and "symmetric". Fractals give structure to complexity, and beauty to chaos. The shapes or time series fill their spaces by fractal geometry. To set the Gaussian random walk has a dimension of 1.5, if the fractal dimension of time series is between 1 and 1.5, which shows that the time series is between a straight line and Gaussian random walk.

2.1.2. Measuring the Fractal Dimension

How jagged the time series is that can be measured by fractal dimension. The entire time series need to be covered by the number of circles which have fixed diameter. If the process keeps doing like this, the relationship between the number of circle and the diameter can be found as follow:

$$N * d^D = 1$$

Where, D = fractal dimension, d = diameter, and N = number of circles.

The fractal dimension can be found by transformed the equation:

$$D = \frac{\log(N)}{\log(\frac{1}{d})}$$

2.2. The Hurst Exponent

In order to measure the smoothness of a time series, the Hurst Exponent is used which related to fractal dimension. Because of the Hurst Exponent is remarkable robust, it is widely used in time series analysis. It can figure out whether the time series are random, a persistent, or an anti-persistent process.The Hurst Exponent, H, was given by H. E. Hurst (1951). The formula is showed as follow:

ISSN: 2688-9323

$$H = \frac{\log(R/S)}{\log(T)}$$

R/S is related to the rescaled range which was develop by Hurst. In 1996 Peters made this method used into the capital markets to find out if there were fractal characteristic and nonlinear behaviors exist in the capital markets. R/S is measured the range of the mean-centered value by divided the T into stand deviation. The T is the duration of the time series.

2.2.1. Rescaled Range Analysis

When analyzing the capital market usually the logarithmic returns are used to time series, because it is more appropriate:

$$x_t = \ln(P_t / P_{t-1})$$

Where x_t is logarithmic return at time t

P_t is price at time t

Begin with the time series,t, with u observations:

$$X_{t,N} = \sum_{u=1}^{t} (e_u - M_N) \tag{1}$$

Where $X_{t,N}$ = cumulative deviation over N periods

e_u = influx in year u

M_N = average e_u over N periods

Taking the difference between the maximum and the minimum levels attained the range (1):

$$R = Max(X_{t,N}) - Min(X_{t,N}) \tag{2}$$

Where R=range of X

Max(X)=maximum value of X

Min(X)=minimum value of X

Hurst divided this range by standard deviation of the original observations which increase with time to compare different types of time series. Then the formula is be got:

$$R/S = (a * N)^H$$

For N is the number of observations and a is a constant, H known as the Hurst exponent.
Taking the logarithmic on both sides:

$$\log(R/S) = \log(a) + H \log(N) \tag{3}$$

Let set if H equal to 0.5, the series is a random walk. So if $H = 0.5$, time series is consistent and independently distributed.. The range $0.5 < H \leq 1$, which means a persistent time series. In the other words, time series have positive correlations. The range $0 < H \leq 0.5$ which implies anti-persistence which means that the time series have negative correlations (E. Peters, 1996).

In the fractal time series, $0.5 < H \leq 1$, time series have persistence and long-term correlations that is different to the, $H = 0.5$, time series are normally distribution.

The Hurst Exponent could shows the time series exist fractal characteristics, but it cannot be measured the susceptibility of a system to sensitive dependence on initial condition which shows the chaotic behaviors. Then an other method that is Largest Lyapunov Exponent, is been selected to test if this time series system is "sensitive dependence on initial condition" and if there exists chaotic characteristics.

2.3. Lyapunov Exponent

The method developed by Wolf et al (1984) let us can calculating the largest Lyapunov exponent, L_1, using experimental data. If L_1 greater than zero would signify that sensitive dependence on initial conditions exists, it also measures stretching in phase space: that is, it measures how rapidly nearby points diverge from one another. Here the definition of phase space is given as following. (Looking at the data, if it is in nonlinear dynamic systems, these problems usually have multiple and perhaps infinite solutions. Many chaotic systems have an infinite number of solutions contain in a finite space. The system is attracted to a region of space, and the set of possible solutions often has a fractal dimension. If all the variables are been known in the system, they can simply plot together on a coordinate system. If there are two variables, variable x and y are been plotted on a standard Cartesian graph, that plot the value of each variable versus the other at the same instant in time. This is called the phase portrait of the system, and it is plotted in phase space. The dimensionality of the phase space depends on the number of variables in the system. The phase space gives us a picture of possibilities in the system). According to Packard et al (1980) and Takens (1981), the method of delays can be used to embed a scalar time series $X(t_i) = (x(t_i+1), x(t_i+2), \cdots x(t_i+n), \cdots)$, into an m-dimensional space as follows:

$$X(t_i) = \{x(t_i), x(t_i+\tau), \cdots, x[t_i+(m-1)\tau)]\}, (i = 1, 2, \cdots)$$

Where m is the embedding dimension (in the other word, this time series has an m-dimensional space) and τ is the delay time factor.

First choosing a point $X(t_0)$, then choosing the second point nearby $X_0(t_0)$, and calculate the distance, L_0, $L_0 = |X(t_0) - X_0(t_0)|$, between these two points. And after a fixed interval of time until t_1, the distance between the two points is measured. If the distance becomes too long, set a ε, if $L_0^{'} = |X(t_1) - X_0(t_1)| > \varepsilon$, keep the $X(t_1)$, and find another point $X_1(t_1)$, let $L_1 = |X(t_1) - X_1(t_1)| < \varepsilon$, this replacement point with an angle of orientation similar to that original point is found. The orientation between the new pair of points should be as close as possible to that of the original pair. Continuing the process until $X(t)$ thought to the end point N. Let the number of iterations be M, then the largest Lyapunov exponent will be:

$$\sigma = \frac{1}{t_M - t_0} \sum_{t=0}^{M} \ln \frac{L_t^{'}}{L_t}$$

This method measure the divergence of nearby points in reconstructed phase space, and indicates how the rate of divergence scales over fixed intervals of time.

For systems where the equation are known, constructing a phase space is simple. Here C-C method is selected to reconstruct a phase space to get delay time and the embedding dimension. C-C method is gave as follow.

2.3.1. Reconstructing a Phase Space Using C-C Method

Kim (1999) proposed C-C method, he developed a technique for choosing either the delay time τ_d or the delay time window τ_w using the correlation integral which introduced by Grassberger and Procaccia (1983).

In order to use it conveniently, the definitions are given as following: t is the index lag, τ_s is sampling time, $\tau_d = t\tau_s$ is the delay time, $\tau_w = (m-1)\tau_d$ is delay time window. $\tau(\tau = t)$ is the value of delay time, and m is embedding dimension, N is size of data set, $M = N-(m-1)\tau$ is the number of embedded points in m dimensional space and defined the $x_i(i=1,2,\cdots M)$ are the points of the reconstruction phase space as

$$X_i = (x_i, x_{i+\tau}, \cdots, x_{i+(m-1)\tau}), X_i \in R^m$$

So, the correlation integral for the embedded time series function is showed in following:

$$C(m,N,r,t) = \frac{2}{M(M-1)} \sum_{1\le i\le j\le M} \theta(r-d_{ij}) \tag{4}$$

For $d_{ij} = \|x(i)-x(j)\|$, $\theta(z) = \begin{cases} 0 & z<0 \\ 1 & z\ge 0 \end{cases}$

As d_{ij} denotes the sup-norm and $r>0$.

Which measure the fraction of the pairs of points for each r and whose sup-norm separation in not great than r.

After the correlation integral was defined, then reconstructed statistic $S(m,N,r,t)$. The formula is given as follow:

$$S(m,N,r,t) = C(m,N,r,t) - C^m(1,N,r,t) ,$$

The time series $\{x_i\}, i=1,2,\ldots,N$ is been subdivided into t disjoint time series. Then $S(m,N,r,t)$ is computed as follows:

$$x^1 = \{x_1, x_{t+1,} \cdots, x_{N-t+1}\}$$
$$x^2 = \{x_2, x_{t+2,} \cdots, x_{N-t+2}\}$$
$$\cdots\cdots\cdots$$
$$x^t = \{x_t, x_{2t,} \cdots, x_N\}$$

For $t=1$, the single time series is $\{x_1, x_2, \cdots, x_N\}$, and

$$S(m,N,r,1) = C(m,N,r,1) - C^m(1,N,r,1)$$

For t=2, the two disjoint time series are $\{x_1, x_3, \cdots, x_{N-1}\}$ and $\{x_2, x_4, \cdots, x_N\}$, each of length $N/2$, and the average of the values of $S(m, N/2, r, 2)$ for these two series:

$$S(m,N,r,2) = \frac{1}{2}\{[C_1(m,\frac{N}{2},r,2) - C_1^m(1,\frac{N}{2},r,2)] + [C_2(m,\frac{N}{2},r,2) - C_2^m(1,\frac{N}{2},r,2)]\}$$

For general t, this becomes

$$S(m,N,r,t) = \frac{1}{t}\sum_{s=1}^{t}[C_s(m,\frac{N}{t},r,t) - C_s^m(1,\frac{N}{t},r,t)]$$

ISSN: 2688-9323

Finally, as $N \to \infty$ we can write

$$S(m,r,t)=\frac{1}{t}\sum_{s=1}^{t}[C_s(m,r,t)-C_s^m(1,r,t)], m=2,3,\dots$$

If the time series data is iid, then for fixed m and t, when $N \to \infty$ for all r, $S(m,r,t)$ equal to 0. But for the real data, things go different: first the data set may be serially correlated, second the data set is finite. These make the $S(m,r,t)\neq 0.$ For that reason, the locally optimal times could be zero crossings of $S(m,r,t)$ or the times at the least variation with r of $S(m,r,t)$. Hence, we select the max and min values of $S(m,r,t)$ of r_j, and the formula is defined

$$\Delta S(m,t)=\max\{S(m,r_j,t)\}-\min\{S(m,r_j,t)\},$$

$\Delta S(m,t)$ measure of the max variation of $S(m,r,t)$ with r. Then the zero crossings of $S(m,r,t)$ and the minims of $\Delta S(m,t)$ are the locally optimal times t. For all m and r, the zero crossings of $S(m,r,t)$ should be nearly the same. For all m, the minims of $\Delta S(m,t)$ should be nearly the same. The first of this locally optimal time is the delay time τ_d .

Appropriate choices for m, N, and r can be found by examining the BDS statistic. The BDS statistic originates from the statistical properties of the correlation integral, and it measures the statistical significance of calculations of the correlation dimension. Even though the BDS statistic cannot be used to distinguish between a nonlinear deterministic system and a nonlinear stochastic system, it is a powerful tool for distinguishing random time series from chaotic or nonlinear stochastic time series. In this paper the BDS statistic is defined as

$$BDS(m,N,r)=\frac{\sqrt{N}}{\hat{\sigma}}[C(m,N,r,t)-C^m(1,N,r,t)],$$

Which converges to a standard normal distribution as $N\to\infty$. Brock et al (1996) had studied this kind of time series. He used three sample sizes, N=100, 500 and 1000, of that time series. After that he used six asymptotic distributions: a bimodal mixture of normal distributions, a double exponential distribution, a student-t distribution with three degrees of freedom, a chi-square distribution with four degrees of freedom, a uniform distribution, and a standard normal distribution to generate by Monte Carlo simulations. The results show that m should be between 2 and 5 and r should be between $\frac{\sigma}{2}$ and 2σ. When $N\geq 500$, the asymptotic distributions were well approximated by finite time series.

σ is mean square error or standard deviation, for using C-C method to calculate the delay time factor and embedding dimension, we let N=3000, $m=2,3,4,5,$ $r_i=\frac{i\sigma}{2}, i=1,2,3,4.$

We then define the following averages of the quantities given by Eqs.

$$\overline{S}(t)=\frac{1}{16}\sum_{m=2}^{5}\sum_{j=1}^{4}S(m,r_j,t),$$

$$\Delta\overline{S}(t)=\frac{1}{4}\sum_{m=2}^{5}\Delta S(m,t),$$

The first zero crossing of $\overline{S}(t)$ or the first local minimum of $\Delta\overline{S}(t)$ to find the first locally optimal time for independence of the data, which gives the delay time τ_d. If we assign equal importance to these two quantities, then we may simply look for the minimum of the quantity.

$$S_{\cot}(t) = \Delta\overline{S}(t) + \left|\overline{S}(t)\right|,$$

This optimal time gives the delay time window $\tau_w = (m-1)\tau_d$.

After reconstruct time series as a phase space from $x_1, x_2, x_3 \cdots$, let τ is delay time factor, and m is embedding dimension, then we get the time series as follow:

$$X(t) = \{x(t_i), x(t_i+\tau), \cdots, x[t_i+(m-1)\tau)]\}, (i = 1,2,\cdots)$$

2.4. Forecasting Models

Then these theories are used to test if the Swedish stock market has fractal and chaotic characteristics. If so the Swedish stock market is a nonlinearity, nonstationary, and chaotic data series, that makes the Swedish stock market extremely complex, chaotic and difficult to predict. However as a investment must minims the risk and maximize the profit, that is the reason a mode is needed to predict more accurately. There are several different approaches on time series forecasting. First there will be linear models including moving average, exponential smoothing,ans Auto-regressive integrates moving average.... Second these are the nonlinear models like the threshold Auto-regressive (TAR) model (Tong 1983), the Auto-regressive conditional heteroscedastic (ARCH) model (Engle 1982), general Auto-regressive conditional heteroscedastic (GARCH) model, chaotic dynamics, and artificial neural networks. The univariate Box-Jenkins (1970) Auto-regressive integrated moving average (ARIMA) model is most classical linear forecasting model and has been widely used for model and forecast financial and economical application. "Although some improvement has been noticed with these nonlinear models, the gain of using them to general forecasting problems is limited." (Zhang 2003) These models can't show other kind of nonlinearity, though they are built for especially type of nonlinearity. Recently the BP neural network models are popularity using in nonlinearity and chaotic system, which are the most widely used model form for time series modeling and forecasting. The reason is that they have flexible nonlinear modeling capability. However, using Bp neural network model to model linear problems has yielded mixed results and may be BP neural network model is over fitting the data, whose adaptively formed based on the feature presented from the data.

The hybrid model is by combining several different models. In this paper, the hybrid model combines ARIMA model and BP neural network model to forecast the capital market trend. From Zhang (2003), there are many empirical studies showed that the hybrid models improved forecasting performance. " Several feed forward neural networks were used to improve time series forecasting accuracy by Pelikan et al (1992) and Ginzburg and Horn (1994). M-competition (1982) in which combination of forecasts from more than one model often leads to improved forecasting performance. The sales forecasting was presented by Luxhoj et al. (1996) using a hybrid econometric and ANN approach. Wedding and Cios (1996) used radial basis function networks and the Box–Jenkins models to describe a combining methodology." Zhang (2003).

Based on the Swedish stock market, Auto-regressive integrated moving average (ARIMA) model, BP neural network and the hybrid model are been chosen. These models are been selected to predict the Swedish stock market and determinate which model is more suitable for Swedish stock market. The "true" or "best" model will be found out.

As follow, ARIMA and BP neural networks' modeling approaches to time series forecasting are reviewed, the hybrid methodology is introduced. The results from the real data set is reported in last section.

2.4.1. The Auto-regressive Integrated Moving Average (ARIMA) Models

Auto-regressive integrated moving average (ARIMA) models have been used in many areas of time series forecasting. Here the ARIMA model defined is given briefly. The function form is given as follow:

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{5}$$

Where y_t represents the actual value and ε_t is the random error. P and q are the number of the Auto-regressive terms and the moving average terms, respectively. $\varphi_i (i = 1, 2, \cdots, p)$ and $\theta_j (j = 1, 2, \cdots, q)$ are parameters of models.

Equation (5) makes up tree parts: one is auto-regression AR(p), another is moving average MA(q) and the third is integration I(d).

When the models are been identified, the timer series need be stationary, after that the partial auto-correlation function (PACF) and the auto-correlation function (ACF) are used to select the ARIMA model. There are also another method to select model, such as the Akaike's information criterion (AIC) and Bayesian information criterion (BIC).

The diagnostic checking of model adequacy is been taken at last step. If ε_t are satisfied, the model is selected.

2.4.2. BP Neutral Network Model

In recent years, due to the development of computer technology and artificial intelligence, as the stock market modeling and forecasting application of new technologies and new methods of providing favorable conditions, artificial neural network, because of its extensive ability to adapt, learning ability and mapping capabilities, achieved in modeling nonlinear multivariable systems amazing achievement. These models are decided by the data and don't need the prior assumption to build the models. The network models maybe widely used in time series for forecasting in the future.

2.4.2.1 BP Neutral Network's Mathematical Model

As follow paper, three layers BP neutral network is took to predict the index of stock market.

In first layer is input layer, that have n nodes, and normalization is done to operate for the original data. Let $x^{'}(t)$ as the original time series data and $x(t)$ as the time series data after normalization.

$$x(t) = \frac{x'(t) - \min\{x'(t)\}}{\max\{x'(t)\} - \min\{x'(t)\}}$$

And the second layer is hide layer. It not only receive the signals from the first layer, but also the delayed signals from itself.

The third layer is output layer. This layer only have one node for output predict value.

The inputs vector: $X = (x_1, x_2 \cdots x_l \cdots x_i)^T$

The hidden layer output vector: $Y = (y_1, y_2 \cdots y_m \cdots y_j)^T$

The output: $Z = (z_1, z_2 \cdots z_n \cdots z_k)^T$

The expected output: $D = (d_1, d_2 \cdots d_n \cdots d_k)^T$

The connection weights between input layer and hidden layer:

$$W = (w_1, w_2 \cdots w_i)^T$$

The connection weights between hidden layer and output layer:

$$V=(v_1,v_2\cdots v_j)^T$$

And the biased and recursion neutral network model has the following mathematical representation.

$$y_m=f(net_m),m=1,2,\cdots,j;net_m=\sum_{m=1}^{j}w_{nm}x_n,n=1,2,\cdots i;$$

$$z_m=f(net_l),l=1,2,\cdots,k;net_l=\sum_{m=1}^{j}w_{ml}y_m,m=1,2,\cdots j;$$

And the hidden layer transfer function often suggests the $f(\cdot)$ logistic function, that is,

$$f(x)=\frac{1}{1+e^{(-x)}},f^{'}(x)=f(x)[1-f(x)]$$

The parameters are estimated such that the cost function of neural network is minimized. Cost function is an overall accuracy criterion such as the following mean squared error:

$$E=\frac{\sum_{i=1}^{N}(t_i-z_i)^2}{N}$$

t_{pi} and z_{pi} are expected output and real output.

2.4.2.2 Learning Part (Network Learning Formula Derivation)

Network learning formula derivation is correcting the connection weights (w_{ij},v_{kj}) and threshold value (θ) and let the cost function decreased along the gradient direction. BP neural network has three layer points: input point x_i, hidden point y_j and output point z_k. The connection weight between input point and hidden point is w_{ij}, the connection weight between hidden point and output point is v_{kj}. When the expected output is d_k, the formula of BP neural network model is should as following:

The hidden point output

$$y_j=f(\sum_i w_{ij}x_i-\theta_j)=f(net_j),net_j=\sum_i w_{ij}x_i-\theta_j$$

The output point

$$z_k=f(\sum_j v_{jk}y_j-\theta_k)=f(net_k),net_k=\sum_j v_{jk}y_j-\theta_k$$

Error function of output point

$$E=\frac{1}{2}\sum_k(x_{i+1}-z_k)^2=\frac{1}{2}\sum_k[x_{i+1}-f(\sum_j v_{jk}y_j-\theta_k)]^2$$

$$=\frac{1}{2}\sum_k\{x_{i+1}-f[\sum_j v_{jk}(\sum_i w_{ij}x_i-\theta_j)-\theta_k]\}^2$$

Then

$$\frac{\partial E}{\partial v_{jk}}=\sum_{m=1}^{k}\frac{\partial E}{\partial z_m}\frac{\partial z_m}{\partial v_{jk}}=\frac{\partial E}{\partial z}\frac{\partial z}{\partial v}$$

ISSN: 2688-9323

And the E is the function of z_m, but there is only one z_k has correlation with v_{jk} and the relation between z_k are independent.

$$\frac{\partial E}{\partial z}=\frac{1}{2}\sum_m -2(t_m-z_m)\frac{\partial z_m}{\partial z_k}=-(d_k-z_k)$$

$$\frac{\partial z_k}{\partial v_{jk}}=\frac{\partial z_k}{\partial net_k}\frac{\partial net_k}{\partial v_{jk}}=f'(net_k)y_i$$

So

$$\frac{\partial E}{\partial v_{jk}}=-(d_k-z_k)f'(net_k)y_j$$

Let δ_k be the error of input point

$$\delta_k=(d_k-z_k)f'(net_k)$$

Then

$$\frac{\partial E}{\partial v_{jk}}=-\delta_k y_j$$

The formula of the hidden point is

$$\frac{\partial E}{\partial w_{ij}}=\sum_k\sum_j\frac{\partial E}{\partial z_k}\frac{\partial z_k}{\partial y_j}\frac{\partial y_j}{\partial w_{ij}}$$

The E is the function of z_k, for one w_{ij}, there exists y_j.

$$\frac{\partial E}{\partial z_k}=\frac{1}{2}\sum_m -2(d_m-z_m)\frac{\partial z_m}{\partial z_k}=-(d_k-z_k)$$

$$\frac{\partial z_k}{\partial y_j}=\frac{\partial z_k}{\partial net_k}\frac{\partial net_k}{\partial y_j}=f'(net_k)v_{jk}$$

$$\frac{\partial y_j}{\partial w_{ij}}=\frac{\partial y_j}{\partial net_j}\frac{\partial net_j}{\partial w_{ij}}=f'(net_j)x_i$$

So

$$\frac{\partial E}{\partial w_{ij}}=-\sum_k(d_k-z_k)f'(net_k)x_i\cdot v_{jk}f'(net_k)=-\sum_k\delta_k f'(net_j)x_i\cdot v_{jk}$$

Let $\delta_j^{'}$ be the error of the hidden point

$$\delta_j^{'}=f'(net_j)\sum_k\delta_k v_{jk}$$

So

$$\frac{\partial E}{\partial w_{ij}}=-\delta_j^{'}x_i$$

Because correcting weights Δv_{jk} and Δw_{ij} are proportional to decrease along the gradient direction, the the formulas of Δv_{jk} and Δw_{ij} are showed as follow:

$$\Delta v_{jk} = -\eta \frac{\partial E}{\partial v_{jk}} = \eta \delta_k y_j$$

$$\delta_k = (d_k - z_k) f'(net_k)$$

$$\Delta w_{ij} = -\eta' \frac{\partial E}{\partial w_{ij}} = -\eta' \delta_j' x_i$$

$$\delta_j' = f'(net_j) \sum_k \delta_k v_{jk}$$

Where the parameter η is called learning rate.

To speed up the learning process, while avoiding the instability of the algorithm, Rumelhart and McClelland (1986) introduced a momentum term ζ , thus obtaining the following learning rule:

$$\Delta w_{ij}(k+1) = -\eta' \frac{\partial E}{\partial w_{ij}} + \xi' \Delta w_{ij}(k)$$

$$\Delta v_{jk}(k+1) = -\eta \frac{\partial E}{\partial v_{jk}} + \xi \Delta v_{jk}(k)$$

The momentum term may also be helpful to prevent the learning process from being trapped into poor local minimize, and is usually chosen in the interval [0,1]. Finally, the estimated model is evaluated using a separate hold-out sample that is not exposed to the training process.

So, the corrected weights are:

$$w_{ij}(k+1) = w_{ij}(k) + \Delta w_{ij}(k+1) = w_{ij}(k) + -\eta \frac{\partial E}{\partial w_{ij}} + \xi \Delta w_{ij}(k)$$

$$v_{jk}(k+1) = v_{jk}(k) + \Delta v_{jk}(k+1) = v_{ij}(k) + -\eta \frac{\partial E}{\partial v_{jk}} + \xi \Delta v_{jk}(k)$$

In this paper C-C method is used to reconstruct a phase space to get delay time and the embedding dimension p which are used to construct the BP neutral network model. The number of the put variables are determinable by embedding dimension, the delay time factors are the delay time factor in node of input layer. This is reasonable, because the phase space contains all the variables in the data and using these variables to construct the BP neural network model and the hybrid model as input variables to predict, can be better explain the Swedish Capital Markets. So the embedding dimension and delay time factors are suggested in this paper to construct the BP neural network model.

2.4.3. The Hybrid Methodology

This hybrid methodology was issued by Zhang (2003). The hybrid model is constructed as following. As so far as we know, ARIMA models maybe good chosen in linear problem, but not work well in the complex nonlinear problem. On the other BP neutral network models are suiting to nonlinear problem, however for the linear problem maybe not suit very well. Hence, the hybrid models combining the ARIMA and the BP neural network model which capture both linear and nonlinear problems.

For the hybrid model, the function form is proposed like

$$y_t = L_t + N_t,$$

Where N_t is the nonlinear part and L_t is the linear part. According to the data, the ARIMA model is selected to model the linear part. The residual e_t only contains the nonlinear parts. The residual is denoted as follow:

$$e_t = y_t - \widehat{L_t}$$

Where $\widehat{L_t}$ is the forecast value from ARIMA model at the time t. After diagnostic checking, the model the appropriately model is build. Then the BP neutral network is used to model the residuals.

$$e_t = f(e_{t-1}, e_{t-2}, \cdots, e_{t-n}) + \varepsilon_t \tag{6}$$

Where ε_t is the random error from the BP neutral network, if the model is an appropriate one. Therefore, the correct model identification is critical. Set (6) as $\widehat{N_t}$ the predict value is defined as follow:

$$\widehat{y_t} = \widehat{L_t} + \widehat{N_t}$$

In the other worlds, the hybrid models are made of two steps. First for the linear parts, the ARIMA model is used to model the linear problem. Second foe the nonlinear part, the BP neutral network is used to model the residuals from the first step. Then the both models are building and the predict values can be gotten.

3. Data

3.1. Database

In this paper, the database are based on the OMX Stockholm 30 Index data and three Swedish companies' stock price, which are Ericsson, Hennes & Mauritz and Nordea Bank are selected. The resource from May 2000 to March 2014. The data is used all the section 3 and all the models.

The OMX Stockholm 30 (OMXS30) is a stock market index for the Stockholm Stock Exchange, which is a capitalization-weighted index that consists of the 30 most-traded stocks. That displays the Swedish capital markets' movements and also shows the trend of the Swedish economic. That why the OMX Stockholm 30 Index data is selected. In order to test these models, the Ericsson, Hennes & Mauritz and Nordea Bank companies' stock are been selected, which take larger parts in constructing the OMX Stockholm 30 Index.

3.2. Output of Hurst Exponent and Lyapounov Exponent

Based on these data, the Hurst analysis are constructed and the Hurst Exponent values are gotten. But the period of the time to calculate the Hurst Exponent maybe have autocorrelation, or correlation, and heteroskedasticity in the error terms in the R/S analysis. In order to overcome autocorrelation, or correlation, and heteroskedasticity, the Newey -West estimator (which was devised by Whitney K.Newey and Kenneth D.West in 1987) is used to improve the ordinary least squares (OLS) regression when the variables have heteroskedasity or autocorrelation (Newey and West 1987). In this paper, the Newey-West estimator is used to calculate the Hurst Exponent. Before using this method, the C-C method is used to get the embedding dimensions and the delayed time.

The Table 1 shows the embedding dimension and the delay time os the OMX Stockholm 30 Index, Ericsson, Hennes & Mauritz B and Nordea.

Table 1. C-C method get the embedding dimensions and the delayed time

C-C	Embedding dimension	Delay time
OMX Stockholm 30 Index	5	1
Ericsson	2	2
Hennes & Mauritz B	3	1
Nordea	2	1

Then the Newey-West estimator is used to calculate the Hurst Exponent, and the Stata soft is used to get the values of the Hurst Exponent.

Table 2. The values of the Hurst Exponent

	Hurst Exponent	T value for Hurst Exponent
OMX Stockholm 30 Index	0.777	20.29
Ericsson	0.654	13.16
Hennes & Mauritz B	0.593	4.03
Nordea	0.716	13.14

The Table 2, the Hurst Exponent values are showed. The T-test is suggested to test if they are difference from 0.5. The null hypothesis $H_0 : H = 0.5$ against the one-sided alternative $H_1 : H > 0.5$. The test statistic is again as given in $t = \frac{\hat{H} - 0.5}{SE_{\hat{H}}} \sim t(n-1)$ and the $SE_{\hat{H}}$ is the Newey-West Standard error. if $t > c$, c is the critical value, the null hypothesis is rejected. In these cases, for n=172 the critical value is around 3.3 in 99.9% interval estimator. All the t value is larger than the critical value, the null hypothesis is rejected, the OMX Stockholm 30 Index, Ericsson, Hennes & Mauritz and Nordea Banks' Hurst Exponent values are significant larger than 0.5. The Lyapounov Exponent values are calculated by Matlab and the values are showed in Table 3.

Table 3. The values of Lyapounov Exponent

	Lyapounov Exponent
OMX Stockholm 30 Index	0.0601
Ericsson	0.1014
Hennes & Mauritz B	0.0427
Nordea	0.2845

From the Table 2 and Table 3, the estimators for the Hurst Exponent for these stock markets are different from 0.5 and the largest Lyapunov exponent of these stock market were been calculated whose values are larger than zero, which show that first the Swedish stock market is clearly fractal and not a random walk. Second what happened yesterday influences what happens today, the Swedish stock markets exist long-term correlations and trends, so the the Swedish stock markets exhibit trend - reinforcing behavior, not mean - trends and the system

ISSN: 2688-9323

shows nonlinear and fractal. Third because of the the Swedish stock markets system is fractal, the return of a time series has a smaller increment of the time will still look the same and has similar statistical characteristics. Finally, the the largest Lyapunov exponent values are larger than zero, the Swedish stock market is sensitive dependence on initial condition, less reliable forecast and has chaotic characteristics. Because of these characteristics, chaos and fractal's theory provides an better explanation for the complex and unstable irregular behavior, unpredictable results regarding deterministic nonlinear system that are sensitive to their initial condition of Swedish stock markets.

3.3. Model Selection

The following, the three models are recommended to forecast the Swedish stock market and compared which model is more appropriate for the Swedish stock markets.

The weekly data of OMX Stockholm 30 Index and three Swedish companies' stock price, which are Ericsson, Hennes & Mauritz and Nordea Bank are selected. From May of 2000 to July 2013 as sample data and from August of 2013 to March of 2014 as predicting data. The plot of these time series data (see Figure 1 to Figure 4) seem they have some economic cycle and in period of they have trend.

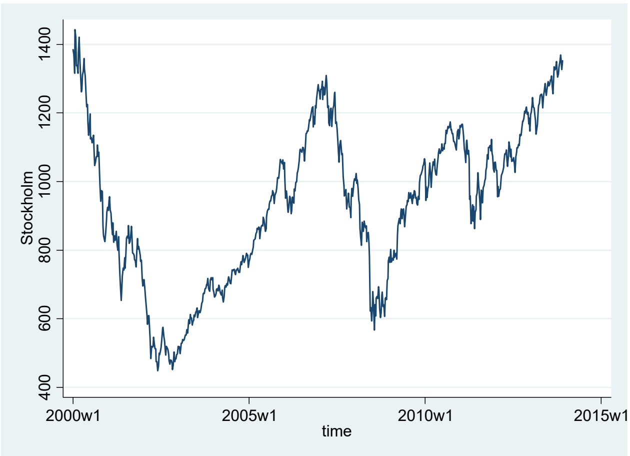

Figure 1. Weekly OMX Stockholm 30 Index price series (2000 - 2014)
Weekly OMX Stockholm 30 Index price series (2000 - 2014)

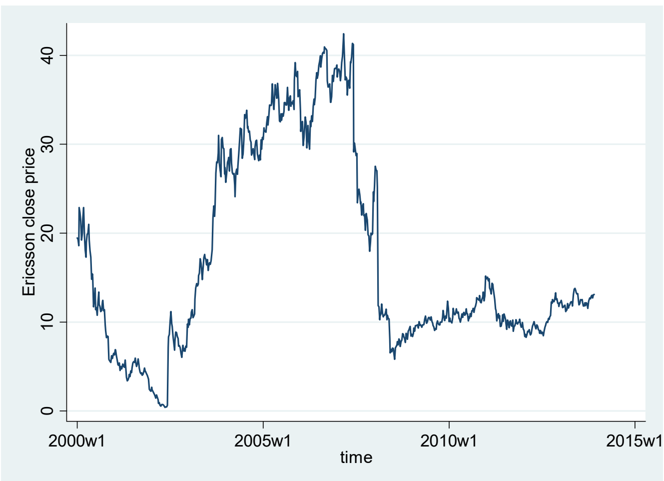

Figure 2. Weekly Ericsson stock price series (2000 - 2014)
Weekly Ericsson stock price series (2000 - 2014)

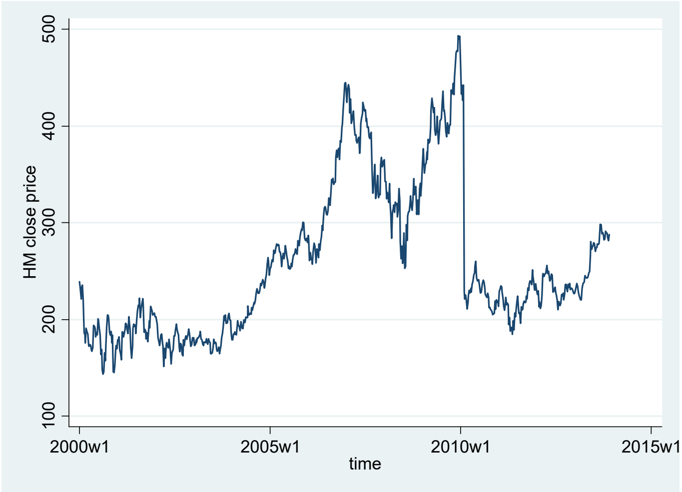

Figure 3. Weekly Hennes & Mauritz stock price series (2000 - 2014)
Weekly Hennes & Mauritz stock price series (2000 - 2014)

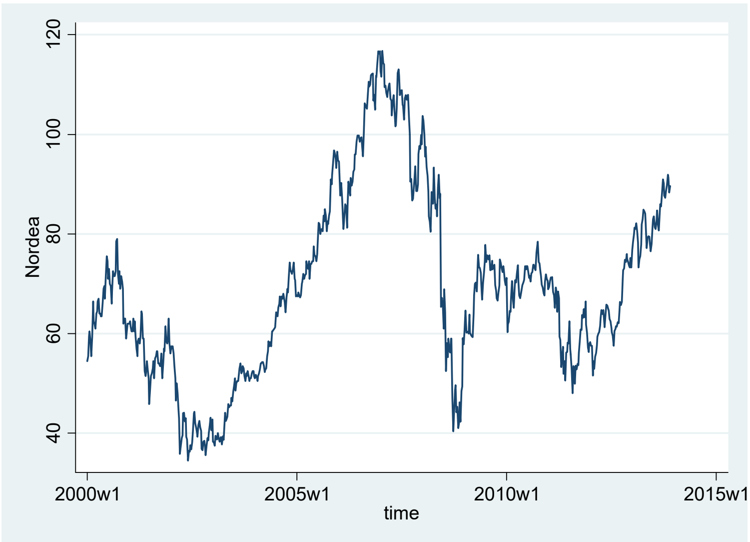

Figure 4. Weekly Nordea stock price series (2000 - 2014).
Weekly Nordea stock price series (2000-2014)

3.3.1. ARIMA Model

In this paper, all ARIMA modelings are implemented via Stata. The train set data is been taken to construct the ARIMA model. The Table 4 shows the construction of the sample data.

Table 4. Sample composition in these data sets

Series	Sample size	Train set (size)	Test set (size)
Stockholm 30 Index	724	2000-2013 (689)	2013-2014 (34)
Ericsson	724	2000-2013 (689)	2013-2014 (34)
Hennes & Mauritz	724	2000-2013 (689)	2013-2014 (34)
Nordea	724	2000-2013 (689)	2013-2014 (34)

Then examined the time series properties of the data using ADF (Augmented Dickey Fuller) test, until the differential function of the close price don't have unite root. The results of the DF test of these time series data are showed in Table.

Table 5. The P-values of ADF (Augmented Dickey Fuller) test.

ADF test	Test statistic	1% critical value	5% critical value	10% critical value	P-value
Stockholm 30 Index	-1.992	-3.430	-2.860	-2.570	0.2900
Ericsson	-1.400	-3.430	-2.860	-2.570	0.582
Hennes & Mauritz	-2.104	-3.430	-2.860	-2.570	0.2430
Nordea	-1.905	-3.430	-2.860	-2.570	0.3296

From the Table 5, the P-values are larger than 0.05, these time series are not stationary, they have unit roots.

In order to obtain the stationary series, the logarithmic and the differential function of the close prices are computed. From the Figure 5 to 8, the data are seemed like the white noise series.

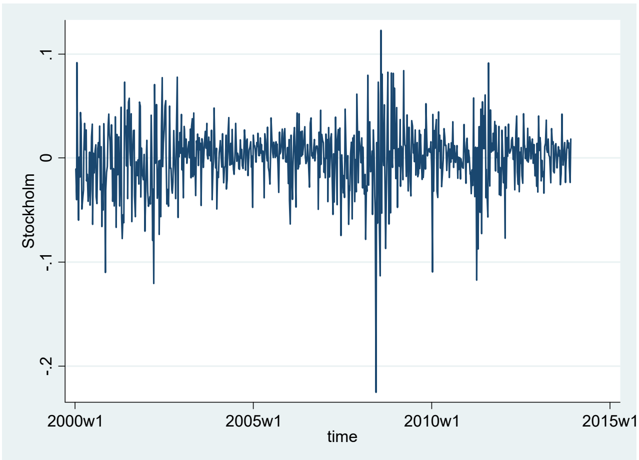

Figure 5. The white noise series of OMX Stockholm 30

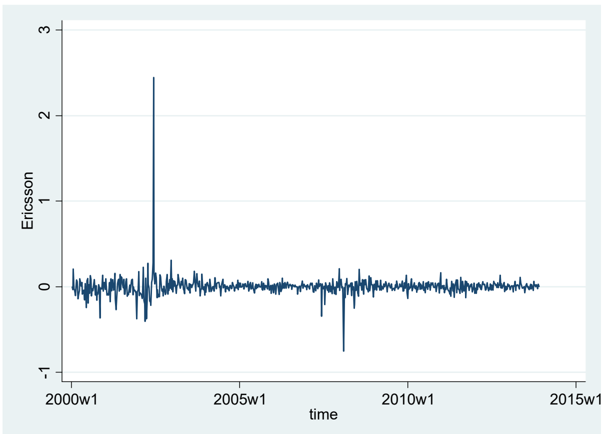

Figure 6. The white noise series of Ericsson stock

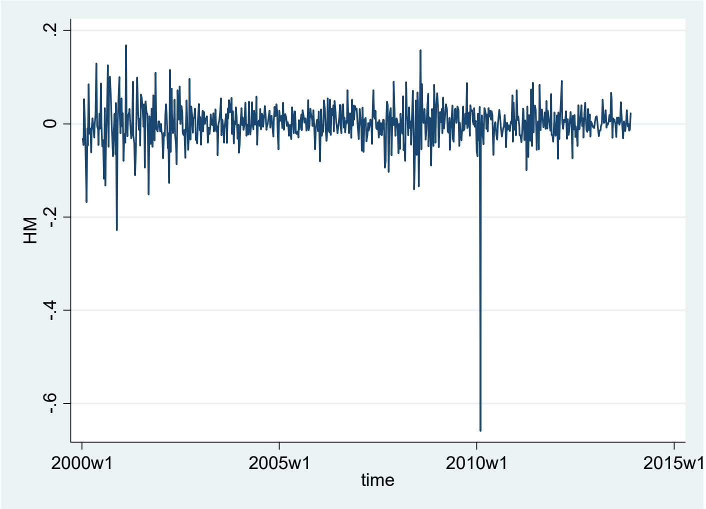

Figure 7. The white noise series of Hennes & Mauritz stock

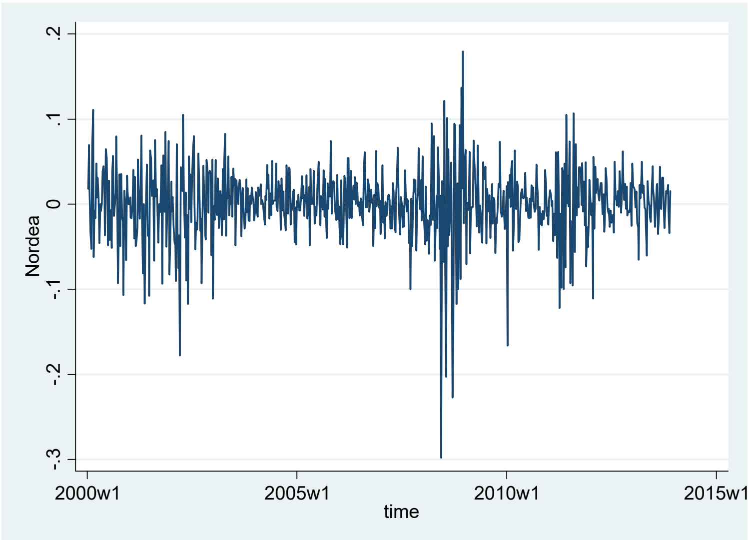

Figure 8. The white noise series of Nordea stock

The results of the ADF test of the logarithmic and the differential function of the close prices are showed in Table 6.

Table 6. The results of the ADF test

ADF test	Test statistic	1% critical value	5% critical value	10% critical value	P-value
Stockholm 30 Index	-27.681	-3.430	-2.860	-2.570	0
Erisson	-23.308	-3.430	-2.860	-2.570	0
Hennes & Mauritz	-27.563	-3.430	-2.860	-2.570	0
Nordea	-28.589	-3.430	-2.860	-2.570	0

From the Table 6, the P-values are showed that the null hypothesis is been reject and these time series are stationary.

After the stationary is identified, the best fit AR parameters and MA parameters should be estimated according to its autocorrelation (AC) function and partial autocorrelation (PAC) function, respectively. The programs are been done in Stata.

The Table 7 shows the ARIMA models of the logarithmic and the differential function of the close price of OMX Stockholm 30 Index, Ericsson, Hennes & Mauritz and Nordea Bank.

Table 7. The ARIMA models of these four stocks

Time series	ARIMA Model	Formula
Stockholm 30 Index	ARIMA(1 4 5 13,1,1)	$z_t = 0.38z_{t-1} - 0.071z_{t-4} + 0.126z_{t-5} + 0.138z_{t-13} + \varepsilon_t$ $-0.413\varepsilon_{t-1}$
Ericsson	ARIMA(1,1,0)	$z_t = 0.117z_{t-1} + \varepsilon_t$
Hennes & Mauritz	ARIMA(1,1,1)	$z_t = 0.868z_{t-1} + \varepsilon_t - 0.908\varepsilon_{t-1}$
Nordea	ARIMA(4,1,4)	$z_t = 0.139z_{t-1} - 0.237z_{t-2} + 0.1z_{t-3} - 0.852z_{t-4} + \varepsilon_t$ $-0.229\varepsilon_{t-1} + 0.281\varepsilon_{t-2} - 0.185\varepsilon_{t-3} + 0.857\varepsilon_{t-4}$

The series z_t was the logarithmic and the differential function returns' function of these time series.

After fitted, ARIMA(p,d,q) has been selected to be the most parsimonious among all ARIMA models. Then the residuals are been tested, if the residuals are white noise series, the fitted model will be found. Once the ultimately fitted model was identified, the equations' form of the model could be obtain.

The Table 8 shows the white noise test of the residuals.

Table 8. The white noise test of the residuals

White noise test	Q Statistic	P-value
Stockholm 30 Index	42.3111	0.3715
Ericsson	42.8726	0.3490
Hennes & Mauritz	30.3507	0.8652
Nordea	28.2553	0.9182

From the Table 8, the residuals are white noise series, so the fitted model will be found.

3.3.2. BP Neural Network Model

A three-layer BP neural network model is developed for the predict of the OMX Stockholm 30 Index, Ericsson, Hennes & Mauritz and Nordea Banks' weekly data set from January 2000 to July 2013 which is used for model training, and the rest data set from August 2013 to March 2014 is used for model verification purpose. From the Table 4, the construction of the data shows clearly. In BP neural modeling process, the input and output data sets for each parameter were normalized to range of [0,1].

The number of neurons in input layer is been calculated by C-C method as equal to embedding dimension and the output layer has been set as 1. In order to determine the optimum number of hidden nodes, a series of different topologies are used. When the training set has the lowest error, the number of the hidden nodes are been settled. The parameters of the network were chosen as follows: the stop criterion of error function was set to 1e-5 and the maximum of the

ISSN: 2688-9323

number of iterations was 5000. Computer program has been performed under MATLAB 7.0 environment. As follows are the Figure for performance of the BP neural network model. The three-layer BP neural network models of the OMX Stockholm 30 Index, Ericsson, Hennes & Mauritz and Nordea Bank are constructed as Table 9.

Table 9. BP model of these four stocks.

BP model	Input nodes	Delay time	Hidden nodes	Output nodes
Stockholm 30 Index	5	1	8	1
Ericsson	2	2	8	1
Hennes & Mauritz	3	1	10	1
Nordea	2	1	4	1

3.3.3. Hybrid Modeling

The proposed algorithm of the hybrid system consisted of two steps. In the first step, to analyze the linear part of the problem, an ARIMA model was employed. In the second step, the residuals from the ARIMA model were modeled by using a neural network model. Since the ARIMA model cannot detect the nonlinear structure of the stock market time series data, the residuals of linear model will contain information about the nonlinearity. The outputs from the neural network can be used as predictions of the error terms of the ARIMA model. The hybrid model utilizes the unique feature and strength of ARIMA model as well as ANN model in determining different patterns. Therefore, it may be favorable to model linear and nonlinear patterns separately by using different models and then combine the predictions to improve the overall modeling and predicting performance. In the hybrid modeling algorithm, the input and output stocks' data sets for each parameter were normalized to the range of [0,1]. In the modeling process, the hybrid model was trained to adjust the model and the number of neurons in input layer was been set by C-C method as equal to embedding dimension, the output layer has been set as 1 and the hidden nodes were depended on the models. The construction of the hybrid model of residuals sets is showed in Table 10 and the models of the hybrid models of OMX Stockholm 30 Index, Ericsson, Hennes & Mauritz and Nordea Bank are constructed as Table 11.

Table 10. Sample composition in residuals sets

Residual's series	Sample size	Train set (size)	Test set (size)
Stockholm 30 Index	724	2000-2013 (689)	2013-2014 (34)
Ericsson	724	2000-2013 (689)	2013-2014 (34)
Hennes & Mauritz	724	2000-2013 (689)	2013-2014 (34)
Nordea	724	2000-2013 (689)	2013-2014 (34)

Table 11. Sample composition in residuals sets of Hybrid model

Hybrid model	Input nodes	Delay time	Hidden nodes	Output nodes
Stockholm 30 Index	5	1	8	1
Ericsson	2	1	4	1
Hennes & Mauritz	3	1	9	1
Nordea	2	1	5	1

3.4. Comparison of Model Performance

In this article, the one-step forecast is suggested and the actual values are been taken to make next step forecast.

The formula is showed:

ISSN: 2688-9323

$$x_{i+1} = g(x_i, \hat{w}_{ij}, \hat{v}_{jk}, \hat{\theta}_j, \hat{\theta}_k)$$

$$g(x_t, \hat{w}_{ij}, \hat{v}_{jk}, \hat{\theta}_j, \hat{\theta}_k) = f\{\sum_j \hat{v}_{jk}[f(\sum_i \hat{w}_{ij} x_i - \hat{\theta}_j)] - \hat{\theta}_k\}$$

The 95% confidence intervals of predict value of these time series could show clearly the performance of the three models. The formula shows as following:

$$\left[\hat{X} - 1.96S, \hat{X} + 1.96S\right]$$

$$S = \sqrt{\frac{\sum(X - \hat{X})^\wedge 2}{n-m}}$$

Where S is the standard variance

X is the actual value

$\hat{X}$ is predict value

$n-m$ is the degree of freedom

m is the number of parameter

n is the number of objective

To evaluate the performance of the forecasting capability, the three evaluation statistics: root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage forecast error (MAPE) to each model are used. They are expressed as below:

$$RMSE = \sqrt{\sum_{i=1}^{n} (x_i - \widehat{x_i})^\wedge 2 / n}$$

$$MAE = \sum_{i=1}^{n} \left|(x_i - \widehat{x_i})\right| / n$$

$$MAPE = \sum_{i=1}^{n} \left|(x_i - \widehat{x_i}) / \widehat{x_i}\right| / n \times 100\%$$

4. Result

4.1. The OMX Stockholm 30 Index

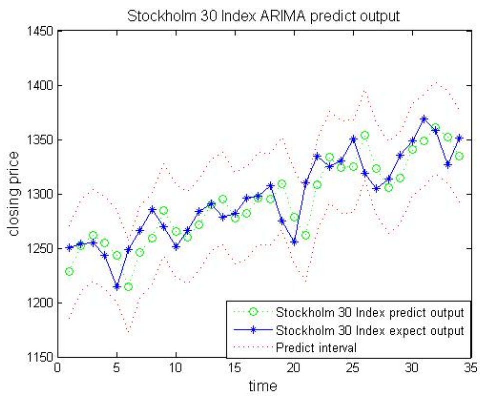

Figure 9. Stockholm 30 index ARIMA predict output

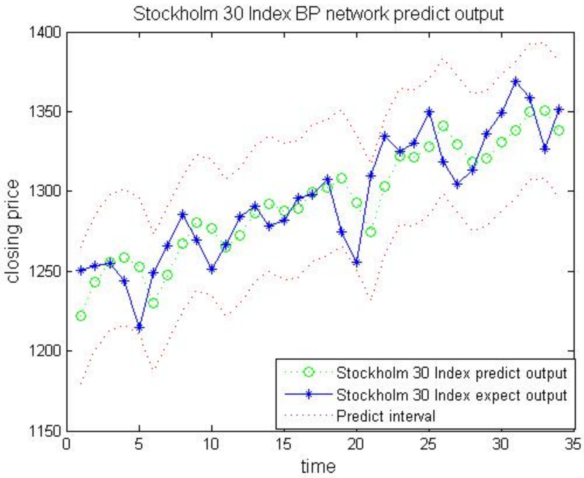

Figure 10. Stockholm 30 index BP network predict output

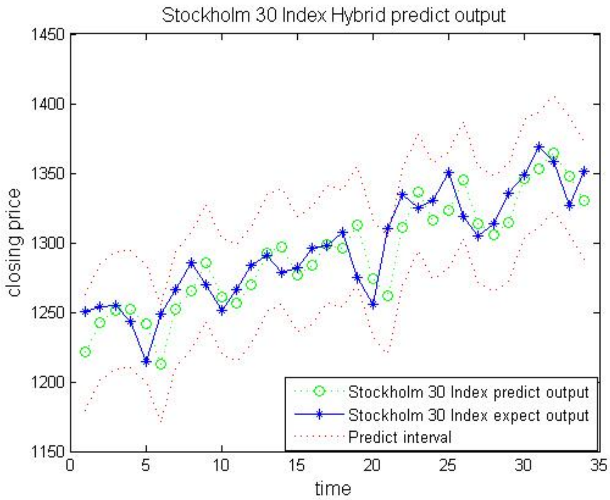

Figure 11. Stockholm 30 index hybrid predict output

Table 12. The OMX Stockholm 30 Index forecasting performance of different model

Stockholm 30 Index	ARIMA	BP neural network	Hybrid model
MAE	16.9639	16.76	16.5558
RMSE	20.3115	20.0892	19.7820
MAPE	1.32%	1.3%	1.29%

From the Figure 9 to the Figure 11, the performance of these three models are been showed, which also give the actual vs forecast values with individual models of ARIMA, BP neural network, and hybrid model. The predict values are closely to the expect values. The predict interval contain all the expect values in 95% confidence interval. From the Table 12, in the OMX Stockholm 30 Index data, the BP neural network model gives slightly better forecasts than the ARIMA model. Applying the hybrid model, the MAE, RMSE and MAPE are less than the ARIMA and BP neural network model, which means the hybrid model is better than the ARIMA and BP neural network model.

4.2. Ericsson

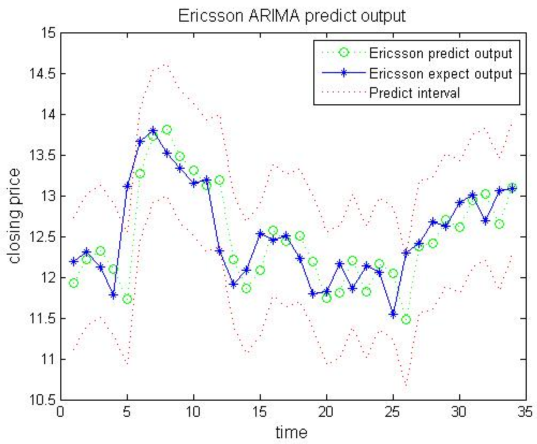

Figure 12. Ericsson ARIMA predict output

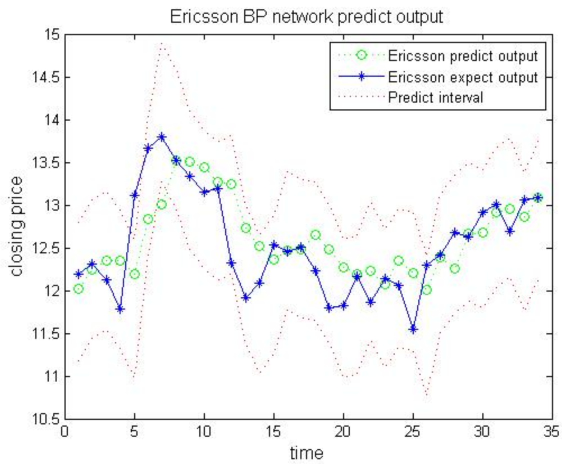

Figure 13. Ericsson BP network predict output

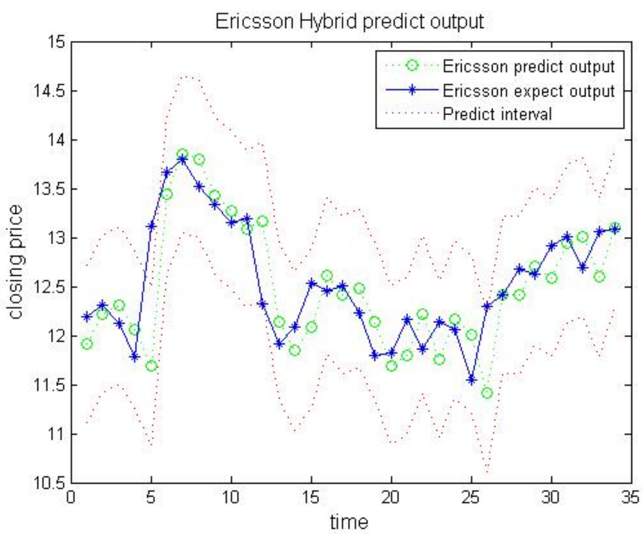

Figure 14. Ericsson hybrid predict output

ISSN: 2688-9323

Table 13. The Ericsson forecasting performance of different model

Ericsson	ARIMA	BP neural network	Hybrid model
MAE	0.2956	0.2937	0.2933
RMSE	0.4010	0.3850	0.3991
MAPE	2.4%	2.38%	2.37%

From the Figure 12 to the Figure 13, the performance of these three models are been showed, which also give the actual vs forecast values with individual models of ARIMA, BP neural network, and hybrid model. The predict values are closely to the expect values. The predict interval contain almost the expect values in 95% confidence interval. From the Table 13, in the Ericsson company's stock data, the BP neural network model gives slightly better forecasts than the ARIMA model. While the hybrid model is better than the ARIMA model, the MAE and MAPE are less than BP neural network model. The hybrid model is better than ARIMA and BP neural network models.

4.3. Hennes & Mauritz

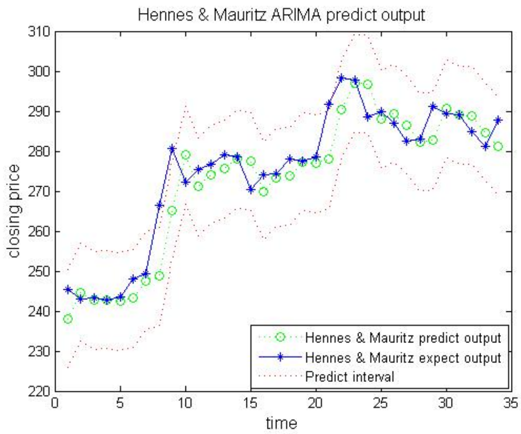

Figure 15. Hennes & Mauritz ARIMA predict output

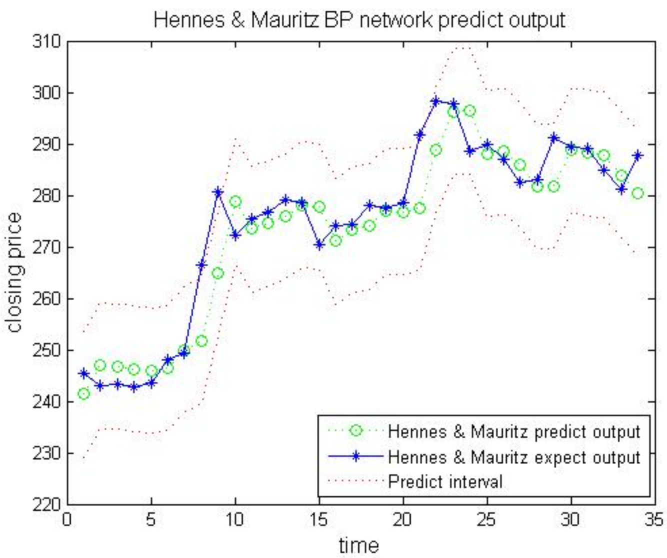

Figure 16. Hennes & Mauritz BP network predict output

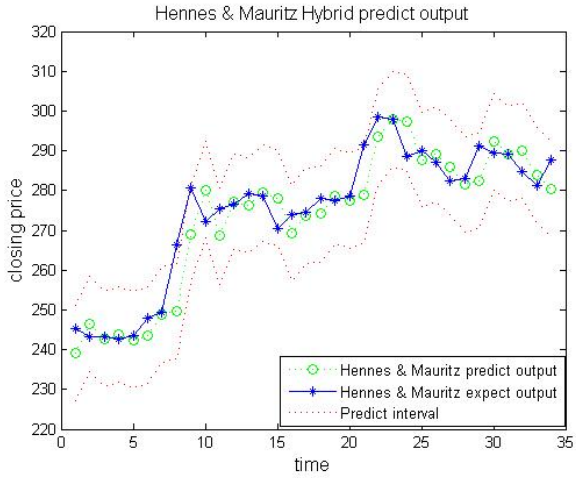

Figure 17. Hennes & Mauritz hybrid predict output

Table 14. The Hennes & Mauritz forecasting performance of different model.

Hennes & Mauritz	ARIMA	BP neural network	Hybrid model
MAE	4.3944	4.3281	4.3158
RMSE	6.1526	5.9477	5.8590
MAPE	1.62%	1.59%	1.59%

From the Figure 15 to the Figure 17, the performance of these three models are been showed, which also give the actual vs forecast values with individual models of ARIMA, BP neural network, and hybrid model. The predict values are closely to the expect values. The predict interval contain almost the expect values in 95% confidence interval. From the Table 14, in the Hennes & Mauritz company's stock market data, the BP neural network model gives a better forecasts than the ARIMA model. Applying the hybrid model, the MAE, RMSE and MAPE are less than ARIMA and BP neural network model respectively.

4.4. Nordea Bank

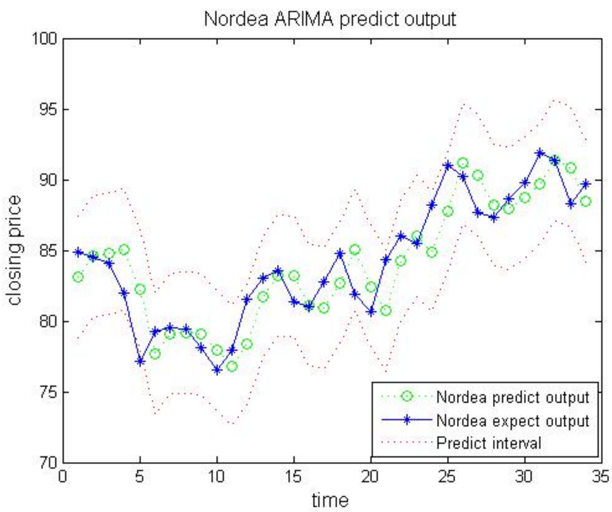

Figure 18. Nordea Bank ARIMA predict output

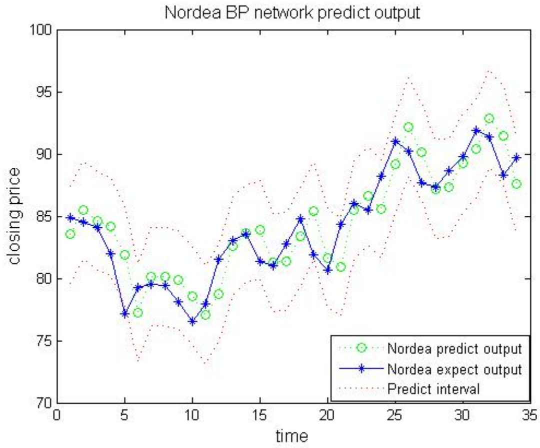

Figure 19. Nordea Bank BP network predict output

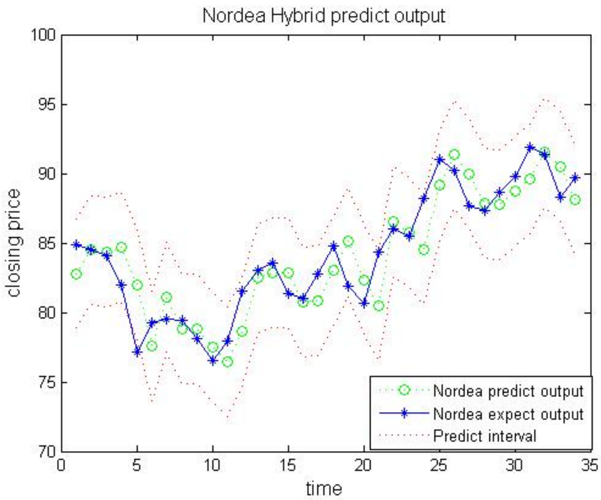

Figure 20. Nordea Bank hybrid predict output

Table 15. The Nordea forecasting performance of different model.

Nordea	ARIMA	BP neural network	Hybrid model
MAE	1.6655	1.6287	1.5723
RMSE	2.0502	1.9511	1.9449
MAPE	1.99%	1.93%	1.88%

From the Figure 18 to the Figure 20, the performance of these three models are been showed, which also give the actual vs forecast values with individual models of ARIMA, BP neural network, and hybrid model. The predict values are closely to the expect values. The predict interval contain all the expect values in 95% confidence interval. From the Table 15, in the Nordea company's stock market data, the BP neural network model gives a better forecasts than the ARIMA model. Applying the hybrid model, the MAE, RMSE and MAPE are less than ARIMA and BP neural network model respectively.

Compared with the predict of the OMX Stockholm 30 Index, Ericsson, Hennes & Mauritz and Nordea Bank, the BP neural network model performances better than ARIMA model in the OMX

ISSN: 2688-9323

Stockholm 30 Index, the Ericsson, Hennes & Mauritz and Nordea bank stock market data, while the results of the hybrid models show that the hybrid models effectively reduce the forecasting error and give a better forecasting of the OMX Stockholm 30 Index, Ericsson, Hennes & Mauritz and Nordea Bank than ARIMA and BP neural network model. This may suggest that neither the ARIMA nor the BP neural network model captures all the patterns in the data and combining two models together can give a better forecasting value of the Swedish capital market.

5. Conclusion

In this paper, the Swedish capital markets have been tested to see if they have fractal and chaotic characteristics, which represent that the Swedish capital markets system are not just purely random walk system. The Fractal Market Hypothesis based on the chaos theory which can provide a better explanation for the Swedish capital markets. The Swedish capital markets represent a state of complex, unstable irregular behavior, unpredictable results regarding deterministic nonlinear systems that are sensitive to their initial condition, which make the Swedish capital markets behavior extremely chaotic and difficult to predict. After comparing all the models, the forecasting performance of each model is assessed by three statistical measures: RMSE, MAE, MAPE. The results of the statistical measures suggest that the hybrid model can be an effective tool to improve the forecasting accuracy obtained by either of the models used separately. Although both the ARIMA and BP neural network models are effective as forecasting models, they could not capture all the patterns of the Swedish capital market and neither can be the best model in every forecasting situation of the time series data. The hybrid model which combined the linear ARIMA and the nonlinear BP neural network model gave better forecasting of the Swedish capital market and the hybrid model can also capture more patterns in the data than the other two models. Then the hybrid model can give a better explanation for the Swedish capital markets.

Based on the characteristics of the capital market, the hybrid model can give a better forecasting than obtained by either of the models used separately. Maybe various combining methods will be found and different models, linear models or nonlinear models, can be combined to improve forecasting accuracy.

References

[1] Edgar E. Peter, Fractal Market Analysis, Economic Science Press 2002.

[2] Sharpe, William F. 1966. “Mutual Fund Performance.” Journal of Business. January, 39, pp. 119–38.

[3] Lintner, John. 1965a. “The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets.” Review of Economics and Statistics. February, 47, pp. 13–37.

[4] Lintner, John. 1965b. “Security Prices, Risk and Maximal Gains from Diversification.” Journal of Finance. December, 20, pp. 587–615.

[5] Mossin, Jan. 1966. “Equilibrium in a Capital Asset Market.” Econometrica. October, 35, pp. 768–83.

[6] Black, F., M. C. Jensen, and M. Scholes (1972): \The Capital Asset Pricing Model: some empirical tests," in Studies in the theory of capital markets, ed. by M. C. Jensen, chap. 2, pp. 79{121. Praeger Publishers.

[7] Ross, Stephen A. 1976. “Arbitrage Theory of Capital Asset Pricing.” Journal of Economic Theory. December, 13, pp. 341–60.

[8] H. E. Hurst, Long-term storage of reservoirs: an experimental study, Transactions of the American society of civil engineers, 116, 1951, 770-799.

[9] B. B. Mandelbrot & J. Van Ness, Fractional Brownian motions, fractional noises and applications, SIAM Review, 10, 1968, 422-437.

[10] R. F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. in Dation, Econometrica 50 (1982) 987–1008.

[11] B. Mandelbrot, The fractal geometry of nature (New York: W. H. Freeman, 1982).

[12] A. Wolf and J. Swift, "Progress in Computing Lyapunov Exponents from Experimental Data", in statistical physics and chaos in Fusion Plasmas, C.W.horton Jr and LE. Reichi eds. (wiley. New York 1984).

[13] N. H. Packard, J.P. Crutchfield, J.D. Farmer, R.S. Shaw, Phys. Rev. Lett. 45 (1980) 712.

[14] F. Takens, in: D. A. Rand, L.S. Young (Eds.), Dynamical Systems and Turbulence, Lecture Notes in Mathematics, vol. 898 Springer, Berlin, 1981, p. 336.

[15] Kim H S, Eykholt R, Salas J D, et al. Nonlinear dynamics, delay times, and embedding windows. Physica D : Nonlinear Phenomena. 1999, 127(1-2) :48-60.

[16] P. Grassberger, I. Procaccia, Physica D 7 (1983) 153.

[17] W. A. Brock, W.D. Dechert, J.A. Scheinkman, B. LeBaron, Econ. Rev. 15 (1996) 197.

[18] G. Peter Zhang."Time series forecasting using a hybrid ARIMA and neural network model". Neurocomputing 2003; 50: 159-175.

[19] G. Zhang, E.B. Patuwo, M.Y. Hu. "Forecasting with artificial neural networks: the state of the art". Int. J. Forecasting 1998; 14:35-62.

[20] Box, G. and Jenkins,G. (1970). Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco.

[21] McClelland, J. L., & Rumelhart, D. E. (1986). Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models. Cambridge, MA: MIT Press.

[22] Newey, Whitney K.; West, Kenneth D. (1987). "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". Econometrica 55 (3): 703–708. JSTOR 1913610.

[23] onathan M.Blackledge, The Fractal Market Hypothesis: Applications to Financial Forecasting, Centre for Advanced Studies, Warsaw University of Technology , vol: 978-83-61993-01-83, pages: 1-105, 2010.

[24] Fei Chen, Francis X. Diebold, A Markov-Switching Multi-Fractal Inter-Trade Duration Model with Application to U.S. Equities.

[25] Kugiumtzis D. State space reconstruction parameters in the analysis of chaotic time series -- the role of the time window length. Physica D: Nonlinear Phenomena. 1996, 95 (1) : 13 - 28.

[26] G. E. P. Box, G. Jenkins, Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, CA, 1970.

[27] H. Tong, Threshold Models in Non-Linear Time Series Analysis, Springer-Verlag, New York, 1983.

[28] J. T. Luxhoj, J. O. Riis, B. Stensballe, A hybrid econometric-neural network modeling approach for sales forecasting, Int. J. Prod. Econ. 43 (1996) 175–192.

[29] S. Makridakis, A. Anderson, R. Carbone, R. Fildes, M. Hibdon, R. Lewandowski, J. Newton, E. Parzen, R. Winkler, The accuracy of extrapolation (time series) methods: results of a forecasting competition, J. Forecasting 1 (1982) 111–153.

[30] E. Pelikan, C. de Groot, D. Wurtz, Power consumption in West-Bohemia: improved forecasts with decorrelating connectionist networks, Neural Network World 2 (1992) 701–712.

[31] D. K. Wedding II, K.J. Cios, Time series forecasting by combining RBF networks, certainty factors, and the Box–Jenkins model, Neurocomputing 10 (1996) 149–168.

[32] I. Ginzburg, D. Horn, Combined neural networks for time series analysis, Adv. Neural Inf. Process. Systems 6 (1994) 224–231.