Prediction of Stock Price of Iranian Petrochemical Industry Using GMDH-Type Neural Network and Genetic algorithm

Meysam Shaverdi\textsuperscript{1,*}, Saeed Fallahi\textsuperscript{2} and Vahhab Bashiri\textsuperscript{1}

\textsuperscript{1}Young Researchers Club, Ilam Branch, Islamic Azad University, Ilam, Iran
\textsuperscript{2}Young Researchers Club, Shiraz Branch, Islamic Azad University, Shiraz, Iran

Abstract

The prediction of stock price is an important task in investment and financial decision-making since stock prices/indices are inherently noisy and non-stationary. In this paper a GMDH type-neural network based on Genetic algorithm is used to predict stock price index of petrochemical industry in Iran. For designing the model, data of seven petrochemical companies is taken from Tehran Stock Exchange (TSE) in decennial range (1999-2008). We instructed GMDH type neural network by 80\% of the experimental data. 20\% of primary data which had been considered for testing the appropriateness of the modeling were entered into the GMDH network. The predicted values were compared with those of experimental values in order to evaluate the performance of the GMDH neural network method. The results obtained by using GMDH type neural network are in excellent agreement with the experimental results and has high performance in stock price prediction.

Keywords: GMDH; Artificial neural networks; Stock price index; Genetic algorithm; Prediction of stock price

\* Corresponding author. e-mail: meisam.shaverdi@gmail.com
1-Introduction

Petrochemical industry as one of the most strategic industries in Iran is faced with challenges that one of the most important of them is investment in. Forty percent of Iran non-petroleum exports is related with Petrochemical industry. The petrochemical industry has ability and opportunity to produce and supply the products with high added value, can play an important role in improving Iran's economic position, eliminate unemployment, job creation and income. Looking at the petrochemical industry share in Iran's economic situation, can be find the real place of this industry in the economy of Iran. Aspects of the industry's performance includes: Create jobs directly and indirectly, the country closer to self-sufficiency in need of some crucial products, improving the economic position of country in the world, the national income through export of petrochemical products, job creation in the middle and downstream units (http://www.naftnews.net). Moreover, Iran is one of the largest manufacturer and exporter of petrochemical producer and the fourth of polyethylene Manufacturer in the world (http://www.forum.boursekala.com).

Stock price prediction is one of the main tasks in all private and institution investors. It is an important issue in investment/ financial decision-making and is currently receiving much attention from the research society. However, it is regarded as one of the most challenging problems due to the fact that natures of stock prices/indices are noisy and non-static (Hall, 1994; Li, Li, Zhu, & Ogihara, 2003; Yaser & Atiya, 1996).

The price changing of stock market is a very dynamic system that has drive from number of disciplines. Two main analytical approaches are fundamental analysis and technical analysis. Fundamental analysis uses the macroeconomics factors data such as interest rates, money supply, inflationary rates, and foreign exchange rates as well as the basic financial status of the company. After scrutiny all these factors, the analyst will then make a decision of selling or buying a stock. A technical analysis is based on the historical financial time-series data. However, financial time series show quite complex data (for example, trends, abrupt changes, and volatility clustering) and such series are often non-stationary, whereby a variable has no clear tendency to move to a fixed value or a linear trend (Cheng & Liu, 2008). The idea of setting up in Iran a well-established stock exchange goes back to the 1930s. In 1968, Tehran Stock Exchange (TSE) established and started trading shares of a limited number of banks, industrial companies and State-backed securities. TSE is a very small exchange compared to all well established Exchanges in terms of the size, turnover, and other indicators; mainly common shares and participation securities only are being traded and there are not any derivatives; nearly impossible to hedge and the risks are very high. In TSE there is a great lack of knowledge and expertise among the TSE’s staff as well as the brokers and investors (Parchehbar Shoghi & Talaneh, 2010).
Several practicable applications about price prediction are demonstrated in previous studies (Abraham, Baikunth, and Mahanti, 2001; Abraham, Philip, and, Saratchandran, 2003; Aiken and Bsat, 1999; Atsalakis & Valavanis, 2009; Baba, Inoue, and Asakawa, 2000; Cao & Parry, 2009; Chang & Liu, 2008; Chang, Liu, Lin, Fan, & Ng, 2009; Chavarnakul & Enke, 2008; Enke & Thawornwong, 2005; Hassan, Nath, & Kirley, 2007; Kim, 2006; Kim and Han, 2000; Kuo, Chen & Hwang, 2001; Tansel et al, 1999; Thammano, 1999; Vellido, Lisboa, & Vaughan, 1999; Yudong & Lenan, 2009; Zhang, Patuwo, & Hu, 1998; Zhu, Wang, Xu, & Li, 2008).

The group method of data handling (GMDH) (Ivakhnenko, 1966) is aimed at identifying the functional structure of a model hidden within the empirical data. The main idea of the GMDH is the use of feed-forward networks based on short-term polynomial transfer functions whose coefficients are obtained using regression combined with emulation of the self-organizing activity behind NN structural learning (Farlow, 1984). The GMDH was developed for complex systems for the modeling, prediction identification, and approximation of multivariate processes, diagnostics, pattern recognition, and clustering in data samples. It has been shown that, for inaccurate, noisy, or small data sets, the GMDH is the best optimal simplified model, with a higher accuracy and a simpler structure than traditional NNs models (Ketabchi, Ghanadzadeh, Ghanadzadeh, Fallahi & Ganji, 2010).

Hwang (2006) used a fuzzy GMDH-type neural network model for prediction of mobile communication. They used input data within a possible extends as; the amount of portion of population, amount of house hold and the amount of average n expenditure per house holds. They showed the proposed neuro-fuzzy GMDH method was excellent for the complicated forecasting problems. Srinivasan (2008) used GMDH network for prediction of energy demand. This paper presented a medium-term energy demand forecasting method that helps utilities identify and forecast energy demand for each of the end-use consumption sector of the energy system, representing residential, industrial, commercial, non-industrial, entertainment and public lighting load. In this study a comparative evaluation of various traditional and neural network-based methods for obtaining the forecast of monthly energy demand was carried out. This paper concluded GMDH very effective and more accurate in producing forecasts than traditional time-series and regression-based models.

The aim of this paper is the application of GMDH type-neural network for prediction of stock price in petrochemical industry. Within this work, we are using financial indices and closing prices in decennial range (1999-2008) that are taken from TSE. The new approach in this paper is using GMDH type-neural network in prediction of stock price for helping investor and financial analyst. The rest of this paper is organized as follows. Section 2 gives literature review of stock price prediction by neural network approach and GMDH methodology. The proposed forecasting model
and the experimental findings from the research is thoroughly described in Section 3. The paper is concluded in Section 4.

2- Research Methodology

2-1- Definition of Stock Price Indices

In this research input data include indices of EPS, PEPS, DPS, P/E and E/P. Stock price is defined as output data. All indices are defined below.

1- Earnings Per Share (EPS). Earnings per share is one of the most important measure of companies strength. The significance of EPS is obvious, as the viability of any business depends on the income it can generate. A money losing business will eventually go bankrupt, so the only way for long term survival is to make money. EPS allows us to compare different companies’ power to make money. The higher the EPS with all else equal, the higher each share should be worth. To calculate this ratio, divide the company’s net income by the number of shares outstanding during the same period (http://investing-school.com).

2- Prediction Earnings Per Share (PEPS). PEPS is the last of Prediction Earnings Per Share. On the other hand, it is unrealized Earnings Per Share.

3- Dividend per share (DPS). DPS is the total dividends paid out over an entire year (including interim dividends but not including special dividends) divided by the number of outstanding ordinary shares issued (investopedia).

4- Price-earnings ratio (P/E). Value investors have long considered the price earnings ratio as one of the single most important numbers available when evaluating a company's stock price. The P/E looks at the relationship between the stock price and the company’s earnings and it is the most popular metric of stock analysis. The price earnings ratio is equal to the price of the stock divided by EPS of common stock (investopedia).

5- Earnings-price ratio (E/P). E/P is a way to help determine a security's stock valuation, that is, the fair value of a stock in a perfect market. It is also a measure of expected, but not realized, growth. It may be used in place of the price-earnings ratio if, say, there are no earnings (as one cannot divide by zero). It is also called the earnings yield or the earnings capitalization ratio. E/P is equal to the EPS of common stock divided by the price of the stock (financial dictionary).

6- Stock price. The Stock price is equal to the last of Stock price which is trading at the one day.
2-2- Group Method of Data Handling (GMDH)

By the GMDH algorithm, a model is represented as a set of neurons in which different pairs of them in each layer are connected through a quadratic polynomial and, thus produce new neurons in the next layer. The formal definition is to find a function, \( \hat{f} \), that can be approximately used instead of the actual one, \( f \), in order to predict output \( \hat{y} \) for a given input vector \( X = (x_1, x_2, x_3, \cdots, x_n) \) as close as possible to its actual output \( y \). Therefore, given number of observations (M) of multi-input, single output data pairs so that

\[
y_i = f(x_{i1}, x_{i2}, x_{i3}, \cdots, x_{in}) \quad (i = 1, 2, 3, \cdots, M).
\]

It is now possible to train a GMDH-type-NN to predict the output values \( \hat{y}_i \) for any given input vector \( X = (x_{i1}, x_{i2}, x_{i3}, \cdots, x_{in}) \),

\[
\hat{y}_i = \hat{f}(x_{i1}, x_{i2}, x_{i3}, \cdots, x_{in}) \quad (i = 1, 2, 3, \cdots, M).
\]

In order to determine a GMDH type-NN, the square of the differences between the actual output and the predicted one

\[
\sum_{i=1}^{M} \left[ \hat{y}_i(x_{i1}, x_{i2}, \cdots, x_{in}) - y_i \right]^2.
\]

is minimized.

By a complicated discrete form of the Volterra functional series (Ivakhnenko, 1966) in the form of

\[
y = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \cdots
\]

which is known as the Kolmogorov-Gabor polynomial (Ivakhnenko, 1966), the connection between the inputs and the output variables can be expressed. As a partial quadratic polynomials consisting of only two variables (neurons) it is in the form of

\[
\hat{y} = G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 + \cdots
\]

Thus such partial quadratic description is recursively used within the network of connected neurons to build the general mathematical relation of the inputs and output variables given in Eq. (4). The coefficients \( a_i \) in Eq. (5) are calculated by the least squares method. It can be seen that a tree of polynomials is constructed using the
quadratic form given in Eq. (5). In other words, the coefficients of each quadratic function $G_i$ are obtained to fit optimally the output in the whole set of input–output data pairs, that is

$$\min_{i} \frac{\sum_{i=1}^{M} (y_i - G_i(i))^2}{M}$$

(6)

In the basic form of the GMDH algorithm, all the possibilities of two independent variables out of the total $n$ input variables are taken in order to construct the regression polynomial in the form of Eq. (5) that best fits the dependent observations $(y_i, i = 1, 2, \ldots, M)$ in a least squares sense (Nariman-Zadeh & Jamali, 2007). The minimum in (6) is $a = (A^T A)^{-1} A^T Y$, where

$$a = \{a_0, a_1, a_2, a_3, a_4, a_5\}$$

(8)

$$Y = \{y_1, y_2, y_3, \ldots, y_M\}^T$$

(9)

$$A = \begin{bmatrix} x_{1p} & x_{1q} & x_{1p} x_{1q} & x_{1p}^2 & x_{1q}^2 \\ x_{2p} & x_{2q} & x_{2p} x_{2q} & x_{2p}^2 & x_{2q}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{Mp} & x_{Mq} & x_{Mp} x_{Mq} & x_{Mp}^2 & x_{Mq}^2 \end{bmatrix}$$

(10)

### 3- The Stock Price Prediction Using the GMDH-type Neural Network

The feed-forward GMDH-type neural network for the stock price was constructed using an experimental data set of seven petrochemical companies from TSE in decennial range (1999-2008). This data set consists of 36 points. For each petrochemical companies, the data was divided into two parts: 80% was used as training data, and 20% was used as test data. The EPS, Prediction PEPS, DPS, P/E and E/P were used as inputs of the GMDH-type network. The Stock prices were used as desired outputs of the neural network.

In order to estimate the stock prices for companies, using the GMDH type network, seven polynomial equations were obtained (Table 1). In this table, $z_1$ is the DPS, and
Prediction of stock price

$z_2$, $z_3$, $z_4$ and $z_5$ are the E/P, P/E, PEPS and EPS, respectively. The proposed model was used to calculate the stock prices (the output data).

In the present study, the stock prices were predicted using GMDH-type NNs. Such a NN identification process needs a suitable optimization method to find the best network architecture. Thus genetic algorithm (GA) is arranged in a new approach to design the whole architecture of the GMDH-type-NNs. It provides the optimal number of neurons in each hidden layer and their connectivity configuration to find the optimal set of appropriate coefficients of quadratic expressions to model stock prices. The best structure in GMDH were reached by two hidden layers with 300 generations, cross over probability of 0.9 and mutation probability of 0.1, to model the stock prices.

In the GMDH architecture, the selection of nodes with the best predictive capability is decided by the GA optimization framework and subsequently the network construction with the corresponding layers are realized based on the search results. For each layer the best node is found based on the objective function. The nodes in the preceding layer connected to the best node in the current layer are marked for realizing the network as search progresses from layer to layer. The developed GMDH neural network was successfully used to obtain seven models for seven companies to calculate their Stock prices. The optimal structures of the developed neural network with 2-hidden layers are shown in Figs.1.
For instance, “ceebbdac” and “abddcceb” are corresponding genome representations for the stock prices of Isfahan and Abadan companies, respectively. In which, $a, b, c, d$ and $e$ stand for (DPS), (E/P), (P/E), (PEPS) and (EPS), respectively. All input variables were accepted by the models. In other words, the GMDH-type-NN provides an automated selection of essential input variables, and builds polynomial equations for the stock prices modeling. These polynomial equations show the quantitative relationship between input and output variables (Table 1).

Table 1  Polynomial Equations of the GMDH Model for the Petrochemical Companies

<table>
<thead>
<tr>
<th>Stock Price of Shiraz company</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_1 = 0.1073 + 6.4957 z_1 + 4.9967 z_2 - 1.1749 z_1^2 + 0.4435 z_2^2 + 10.7362 z_1 z_2$</td>
</tr>
<tr>
<td>$Y_2 = 798.3493 - 1.5289 z_2 + 3.7162 z_3 + 0.4897 z_2^2 - 0.0278 z_3^2 + 0.2282 z_2 z_3$</td>
</tr>
<tr>
<td>$Y_3 = 0.1073 + 6.4957 z_1 + 4.9967 z_2 - 1.1749 z_1^2 + 0.4435 z_2^2 + 10.7362 z_1 z_2$</td>
</tr>
<tr>
<td>$Y_4 = 0.0006 + 2.0071 Y_3 - 0.8481 Y_4 - 0.0027 Y_3^2 - 0.0014 Y_4^2 + 0.0040 Y_3 Y_4$</td>
</tr>
<tr>
<td>$Y_5 = -2.2330 + 0.1003 Y_3 + 98.3436 z_2 + 0.0012 Y_3^2 - 3.1300 z_2^2 - 0.1226 Y_3 z_2$</td>
</tr>
</tbody>
</table>
Stock Price of Isfahan company

$Y_1 = -134.7807 + 27.2696 z_1 - 1548.2494 z_2 + 0.0026 z_1^2 + 53.2196 z_2^2 - 0.5187 z_1 z_2$

$Y_2 = 0.0008 + 0.004 z_1 + 0.0004 z_2^2 - 0.000005 Y_2^2 + 0.0290 z_1 Y_2$

Value = 0.00001 + 1.5189 Y_1 - 0.4950 Y_2 + 0.000008 Y_1^2 + 0.00005 Y_2^2 - 0.00006 Y_1 Y_2$

Stock Price of Farabi company

$Y_1 = -68.3041 - 641.5366 z_5 + 23.1970 z_2 + 22.2012 z_5^2 - 0.0034 z_2^2 - 0.5165 z_1 z_5$

$Y_2 = 25.0939 + 1989.5347 z_4 + 3.0584 z_4 + 14.7334 z_5^2 - 0.0012 z_2 + 0.5544 z_3 z_4$

Value = -1.0988 + 1.1945 Y_1 - 147.4012 z_3 + 0.00009 Y_1 - 0.2267 Y_3 z_5$

Stock Price of Arak company

$Y_1 = - 0.0119 - 9.3280 z_1 +11.3407 z_2 + 0.0121 z_1^2 + 0.0022 z_3^2 - 0.0107 z_1 z_2$

$Y_2 = 0.0043 + 2.8490 z_4 + 4.8269 z_5 + 0.00020 z_2^2 - 0.00003 z_1 z_5$

Value = 0.0001 - 0.8514 Y_1 - 0.00001 Y_1^2 - 0.00001 Y_1 Y_2$

Stock Price of Khark company

$Y_1 = 1.8217 - 122.3323 z_1 + 122.9404 z_5 - 0.2778 z_1^2 - 0.2789 z_5^2 + 0.5567 z_1 z_5$

$Y_2 = -801.5443 + 429.6301 z_3 + 3.2927 z_4 + 6.6109 z_3^2 - 0.0053 z_1 + 0.3271 z_1 z_3$

Value = 0.0006 - 0.3223 Y_3 + 1.1297 Y_4 - 0.00002 Y_4^2 - 0.00001 Y_3 Y_4$

Stock Price of Sarmayeh company

$Y_1 = -3874.9378 + 30.6341 z_1 + 4.5220 z_4 - 0.0449 z_1^2 + 0.0039 z_4^2 - 0.0010 z_1 z_4$

$Y_2 = -801.5443 + 429.6301 z_3 + 3.2927 z_4 + 6.6109 z_3^2 - 0.0053 z_1 + 0.3271 z_1 z_3$

Value = 0.0006 - 0.3223 Y_3 + 1.1297 Y_4 - 0.00002 Y_4^2 - 0.00001 Y_3 Y_4$

Stock Price of Abadan company

$Y_1 = 16.3105 + 663.3362 z_3 - 0.3005 z_3 + 4.3702 z_3^2 + 0.0013 z_3^2 + 0.1927 z_1 z_3$

$Y_2 = 0.0001 + 0.6750 Y_3 + 0.3700 Y_2 - 0.0000009 Y_2^2 - 0.00003 Y_3 Y_2$

Value = 0.0003 + 1.0106 Y_4 - 0.0058 Y_6 + 0.00008 Y_3^2 + 0.0001 Y_4^2 - 0.0002 Y_1 Y_6
Our proposed models behavior in prediction of the stock prices is demonstrated in Figs. 2. The results of the developed models give a close agreement between observed and predicted values of the stock prices.
Prediction of stock price

Fig. 2 Plot of the stock price against data set number to illustrate the prediction ability of the GMDH model; (○) Experimental Points; (+) Calculated Points. a) Shiraz, b) Isfahan, c) Khark, d) Farabi, e) Arak, f) Abadan, g) Sarmayeh

In order to determine the accuracy of the models some statistical measures are given in Table 2.

Table 2 Model Statistics and Information for the GMDH-Type NN Model for the Prediction of Stock Price.

<table>
<thead>
<tr>
<th>Company</th>
<th>Set</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shiraz</td>
<td>Training</td>
<td>0.99</td>
<td>110</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>0.99</td>
<td>52</td>
<td>43</td>
</tr>
<tr>
<td>Khark</td>
<td>Training</td>
<td>0.98</td>
<td>1696</td>
<td>1316</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>0.99</td>
<td>1069</td>
<td>863</td>
</tr>
<tr>
<td>Abadan</td>
<td>Training</td>
<td>0.99</td>
<td>876</td>
<td>663</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>0.98</td>
<td>1667</td>
<td>1235</td>
</tr>
<tr>
<td>Farabi</td>
<td>Training</td>
<td>0.98</td>
<td>1135</td>
<td>979</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>0.99</td>
<td>1176</td>
<td>940</td>
</tr>
<tr>
<td>Arak</td>
<td>Training</td>
<td>0.98</td>
<td>530</td>
<td>427</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>0.99</td>
<td>381</td>
<td>295</td>
</tr>
<tr>
<td>Sarmayeh</td>
<td>Training</td>
<td>0.98</td>
<td>3082</td>
<td>1958</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>0.99</td>
<td>1460</td>
<td>1158</td>
</tr>
</tbody>
</table>
These statistical values are based on $R^2$ as absolute fraction of variance, RMSE as root-mean squared error, and MAD as mean absolute deviation, which are defined as follows:

$$R^2 = 1 - \left[ \frac{\sum_{i=0}^{M} (Y_{i(model)} - Y_{i(actual)})^2}{\sum_{i=0}^{M} (Y_{i(actual)})^2} \right],$$

$$RMSE = \left[ \frac{\sum_{i=0}^{M} (Y_{i(model)} - Y_{i(actual)})^2}{M} \right]^{1/2},$$

$$MAD = \left[ \frac{\sum_{i=0}^{M} |Y_{i(model)} - Y_{i(actual)}|}{M} \right].$$

4- Conclusion

Nowadays petrochemical industry due to the increase in infrastructure investment, is an attractive market for investment. It is supposed that petrochemical materials production will rise according to the increase in the petroleum demand. Thus modeling a soft computing system for stock price prediction in petrochemical industry is important for or all traders and financial consultant to decrease their investment risk and increase the profit of stockholders.

In this study, a feed-forward GMDH-type NN model is developed using experimental data set from Tehran Stock Exchange. In this work, the daily stock price of seven premier petrochemical companies in decennial range (1999-2008) is massed. The Stock prices were estimated by the GMDH model and the results compared with the experimental data. Despite the complexity of the system studied, the GMDH model result to a good prediction. The agreements between the experimental and calculated data were found to be excellent. Comparison of the other stock price prediction methods like ANN, Double moving average, Single and Double exponential smoothing and time series model are left for future research.

References

Prediction of stock price

[18] Investopedia,. www.investopedia.com


Received: June, 2011