



Prediction of Stock Price using Particle Swarm Optimization Algorithm and Box-Jenkins Time Series

Shokrolah Khajavi

Professor of Accounting, Shiraz University, Shiraz, Iran
(Corresponding Author)
shkhajavi@rose.shirazu.ac.ir

Fateme Sadat Amiri

Msc of Accounting, Shiraz University, Shiraz, Iran
fa.s.amiry@gmail.com

ABSTRACT

The purpose of this research is predicting the stock prices using the Particle Swarm Optimization Algorithm and Box-Jenkins method. In this way, the information of 165 corporations is collected from 2001 to 2016. Then, this research considers price to earnings per share and earnings per share as main variables. The relevant regression equation was created using two variables of earnings per share and price to earnings per share, and stock prices were predicted through particle swarm optimization algorithm in MATLAB. IBM SPSS was used to predict stock prices with Box-Jenkins time series. The Results indicate that particle swarm optimization algorithm with 4% error and Box-Jenkins time series with 19% error, have the potential to predict stock prices of companies. Moreover, PSO algorithm model predict stock prices more precisely than Box-Jenkins time series. Also by using EViews 7 software, the results of Wilcoxon-Mann Whitney statistics showed that PSO algorithm predicts the stock price more accurately.

Keywords:

Box-Jenkins Time Series, Earnings per Share, Particle Swarm Optimization (PSO) Algorithms, Price to Earnings Ratio, Stock Price.



1. Introduction

In capital markets have numerous investors that want to obtain more wealth. Some reach to their goals and some loss their capital. So, become familiar with deferent aspects of Tehran Stock Exchange help to predict the future of it and its changes, decrease investment risk and/or yield more return (Mehrani and Mehrani, 2003). In practice, Presence of a dynamic and non-linear system in regard to market behaviors would challenge available models (Alborzi et al, 2008). When conditions and limitations of real world are considered, optimization issue would not be easily solved by mathematical methods (Sharifi et al, 2011). It needs smart and developed tools such as Particle Swarm Optimization (PSO). This tool as a smart system can detect non-linear relation between incomes and outcomes according to total data, and recognize fundamental relations among them (Alborzi et al, 2008). So, the research's issue is that whether we can predict corporations' stock price using Particle Swarm Optimization (PSO) and predict stock price using Box-Jenkins time series. In addition, Particle Swarm Optimization and Box-Jenkins time series lead to similar results or not.

2. Literature Review

2.1. Particle Swarm Optimization Algorithm (PSO)

Primary idea of PSO algorithm was expressed by Eberhart, computer scientist, and Kenedy, psychologist of social issues in 1995 (Kenedy and Eberhart, 1995). Since then, it has been widely used to solve a broad range of optimization problems. It attempts to mimic the natural process of group communication to share individual knowledge when such swarms flock, migrate, or hunt. If one member sees a desirable path to go, the rest of this swarm will follow quickly (Karegowda and Kumari, 2013). Particle swarm optimization (PSO) is an evolutionary computation technique, and its concept originated as a simulation of a simplified social system. PSO is distinctly different from other evolutionary-type methods in a way that it does not need complex encoding and decoding processes and does not use the operation and it takes real numbers as particles in the aspects of representation solution. PSO system combines local search method (through self-experience) with global search methods (through neighboring experience),

attempting to balance exploration. PSO search uses a population (called swarm) of individuals (called particles) that are updated from iteration to iteration. Each particle represents a candidate position (i.e., solution) to the problem at hand. A particle is treated as a point in an M -dimension space, and the status of a particle is characterized by its position and velocity. Initialized with a swarm of random particles, PSO is achieved through particle flying along the trajectory that will be adjusted based on the best experience or position of the one particle (called local best) and ever found by all particles (called global best). PSO updates a population of particles with the internal velocity and attempts to profit from the discoveries of themselves and previous experiences of all other companions.

In every search iteration, each particle is updated by following two "best" values. The first one is the best solution (fitness) it has achieved so far, and this value is called $pBest$. Another "best" value obtained so far by any particle in the population, and this best value is a global best and called $gBest$. After finding the two best values, the particle updates its velocity and positions with following formulas:

$$v_{id}(k+1) = wv_{id}(k) + c_1r_1(p_{id}(k) - x_{id}(k)) + c_2r_2(p_{gd}(k) - x_{id}(k)) \quad (1)$$

$$x_{id}(k+1) = x_{id}(k) + v_{id}(k+1) \quad (i = 1, 2, \dots, m; d = 1, 2, \dots, D) \quad (2)$$

In (1), $P_i = (p_{i1}, p_{i2}, \dots, p_{iD})$ is the best previous position of the i th particle (also known as $pBest$). According to the different definitions, there are two different versions of PSO, with local version, and particles only have information of their own and their neighbors' bests, rather than that of the entire neighborhood. If $P_g = (p_1, p_2, \dots, p_D)$ is the best position among all the particles in the swarm (also known as $gBest$), such a version is called the global version. If P_g is taken from some smaller number of adjacent particles of the population (also known as $lBest$), such a version is called the local version. In (1) and (2) k represents the iterative number and D is dimension of particles, and range of particles are all determined by the problem to be optimized. Variables c_1, c_2 are learning factors, usually $c_1 = c_2 = 2$, which represent the weighting of the stochastic acceleration terms that pull each particle towards $pBest$ and $gBest$

positions. Thus, adjustment of these constants changes the amount of “tension” in the system. $r_1 \sim (0, 1)$, $r_2 \sim (0, 1)$, and w are an inertia weight, which is initialized typically in the range of $[0, 1]$. Inertia weight controls the impact of previous historical values of particle velocity on its current one. A larger inertia weight pressures towards global exploration (searching new area), while a smaller inertia weight pressures toward fine tuning the current search area. Particle’s velocities on each dimension are confined to a maximum velocity $V_{\max} \sim x$ which is a parameter specified by the user. If the iteration formulas would cause the velocity on that dimension to exceed $V_{\max} \sim x$, then the velocity on that dimension is limited to V_{\max} . The termination criterion for the iterations is determined according to whether the max generation or a designated value of the fitness of P_g is reached (Lin et al, 2014).

2.2. Box-Jenkins model

Box-Jenkins is a 3 stage repeat procedure. First stage, is trial recognition stage which is performed using auto-correlation and partial auto-correlation functions. In the second stage or estimation stage, model parameters are estimated and, third stage is called stage of vector precision determination stage, in which adequacy of trial recognition and estimation of model are analyzed using statistical tests such as independency of model error values. When final model was obtained after adjustments and modifications, it is used for prediction of future values of time series.

2.3. Research hypothesis

- 1) Companies` stock price can be predicted using Particle Swarm Optimization Algorithm.
- 2) Companies` stock price can be predicted using Box-Jenkins time series.
- 3) Particle Swarm Optimization Algorithm and Box-Jenkins time series lead to similar results.

PSO and Box-Jenkins time series have been used to predict stock price. PSO algorithm has been implemented in MATLAB, and Box-Jenkins time series has been implemented in IBM SPSS 21 application. Third hypothesis would be tested using Wilcoxon-Mann Whitney test. Kolmogorov-Smirnov test with assurance level of 95% has been used to review if data is normal.

3. Methodology

In order to implement algorithm and answer to research hypotheses Oliveira (2013) has been used (Oliveira et al, 2013). This is an after-event and semi-experimental research. Regarding stock price, in first step information of all companies listed in Tehran share exchange was extracted using Tadbir application, since for many companies financial information of 16 years was not available, they were removed from research population. Finally, a number of 165 companies were selected as sample. Due to incompleteness of financial information of majority of companies in 16 consecutive years, two factors of estimated earnings per share and price to earnings per share were used as effective factors on companies` stock price. The following limitations have been considered when choosing sample:

- 1) Company`s stock price are available for 16 years.
- 2) Companies have announced their annual estimated earnings and price to earnings per share for all 16 years.
- 3) They are not financial intermediates or banks.

3.1. Research variables

Earnings per share (EPS) is calculated from dividing earnings, after tax deduction, to total shares, and indicates the amount a company has earned in a certain period for per normal share.

Price to earnings per share (P/E) includes price ratio to the earnings of per share, which indicates the period needed for capital return from future earnings of share.

Earnings per share reported by enrolled corporations have been extracted from Tadbir application, and it is calculated as shown in equation 4:

$$\frac{\text{net earnings (losses)}}{\text{number of stocks}} = \text{earnings (losses) per shares}$$

This factor plays a main role in decisions of investors about share purchase and decisions of shareholders about selling or holding share (Jahankhani and Safarian, 2003).

Share price to share earnings ratio is calculated using stock prices at the end of period and earnings per share through equation 5:

$$\frac{\text{price per stock}}{\text{earnings (losses) per stock}} = \text{price ratio to earnings (losses) per share}$$

In this research, we want to predict dependent variable of stock price using two independent variables including earning per share, and price to earnings per share.

3.2. Research model

To obtain level of relation between earnings per share and price to earnings per share, multi-linear regression has been used. Multi-linear regression equation is shown below:

Equation 6:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + c$$

In which:

X: between 1 and n, independent variables

β : independent variable coefficients,

C: error amount,

Y: dependent variable.

Simulation of prediction function of bankruptcy according to effective variables (earnings per share and price to earnings per share) has been modeled using equation 7.

Equation 7:

$$Price = a EPS + b P/E + c$$

In this equation, Price indicates dependent variable of stock price which can be determined using independent variables of earnings per share (EPS) and price to earnings per share. (a) indicates dependence coefficient of earnings per share and stock price, (b) indicates dependence coefficient of price to earnings per share and stock price, and (c) is considered as error amount.

To solve this equation by PSO algorithm using data related to 165 companies being reviewed, data related to 140 companies has been used to learn the algorithm, and data related to 25 companies has been used to test PSO algorithm. So, equation 8 is formed:

Equation 8:

$$\begin{bmatrix} \text{price 1} \\ \text{price 2} \\ \vdots \\ \text{price n} \end{bmatrix} = \begin{bmatrix} 1 & \text{EPS 1} & \frac{P}{E} 1 \\ 1 & \text{EPS 2} & \frac{P}{E} 2 \\ \vdots & \vdots & \vdots \\ 1 & \text{EPS n} & \frac{P}{E} n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + C$$

Using linear regression, a, b, c values are obtained as equation 9.

Equation 9:

$$6.2\% < a < 7.3\%$$

$$24.2\% < b < 25\%$$

$$35\% < c < 51.3\%$$

In implementation of PSO algorithm, outcome values are in this range. These ranges are obtained through PSO algorithm in MATLAB.

4. Results

This equation expresses that almost 6.2% to 7.3% of stock price is under influence of earnings per share, about 24.2% to 25% is under influence of price to earnings per share, and about 35% to 51.3% is under other factors. Considering the presence of values related to earnings per share and price to earnings per share for all companies in 16 years, replaced amount of earnings per share and price to earnings per share are considered equal to average of 16 years values in equation. Stock prices of companies can be predicted through solving this equation and determining earnings per share and price to earnings per share using PSO algorithm with 4% error.

Fig1 is drawn considering equation 9 and prediction of stock price using PSO algorithm.

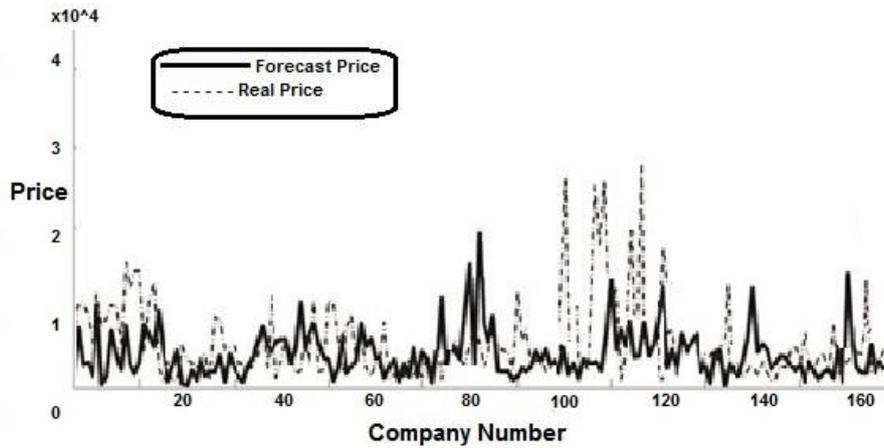


Fig 1: actual and predicted stock price using PSO algorithm

In fig 1, actual price of shares in 16 years period (2001-2016) is shown in dotted line and predicted price for 16 years (2001-2016) period is shown in bold line. Horizontal line indicates the number of corporations and vertical line indicates stock prices. The interesting point is that PSO algorithm with 4% of error can predict stock prices. Through review of great economic events (Nobel, 2008), the reason behind the sudden rise in actual price of shares and its non-predictability is some points may be explained this way that companies located in these points include sugar, steel, concrete, and fireproof products

companies, that their stock prices have become unpredictable in research time span due to sudden changes in global prices of these products and, subsequently high volume of supply and demand. According to what was mentioned about factors creating difference between predicted and actual prices, it can be concluded that PSO algorithm with 4% error can predict stock prices. In fig 2, predicted and actual stock prices have been indicated using Box-Jenkins time series, and in fig 3, comparison of residual ACF and ACF functions of Box-Jenkins have been indicated.

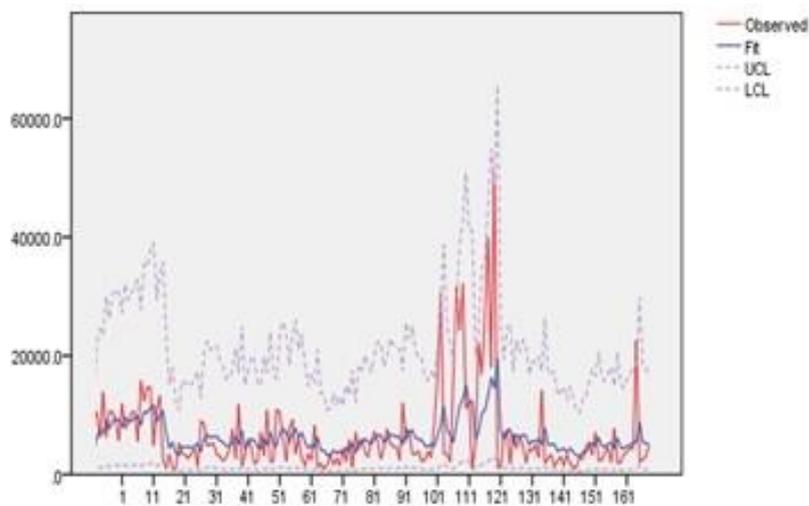


Fig 2: actual and predicted stock prices using Box-Jenkins time series

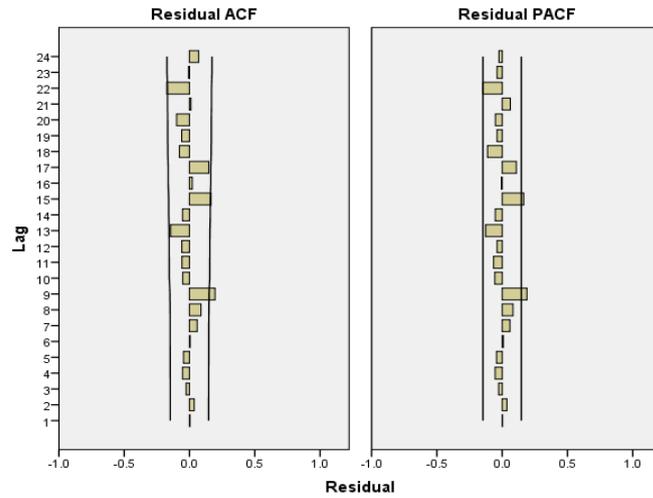


Fig 3: comparison of residual ACF and PACF functions of Box-Jenkins model

According to hypothesis 2, Box-Jenkins time series with 19% error can predict stock prices. The results indicate that both PSO algorithm and Box-Jenkins time series have the potential to predict stock prices; but they do not lead to similar results.

Since distribution of error values of both models of PSO algorithm and Box-Jenkins time series are not normal, and since number of observations is low (30), Wilcoxon-Mann Whitney test was used to review the third hypothesis. According to table 1, significance of

Wilcoxon-Mann Whitney test indicates that in assurance level of 95%, error in prediction of PSO algorithm model is significantly lower than error of Box-Jenkins time series model. So, PSO algorithm can predict stock prices more precisely.

Moreover, table 2 shows comparison information in order to review the results of prediction of stock prices with PSO algorithm and Box-Jenkins time series.

Table 1: results of Wilcoxon-Mann Whitney test

Average of residual of PSO algorithm model	Average of residual of Box-Jenkins time series model	Z test	Significance level
21	57	2.40	0.02

Table 2: comparison of stock prices with PSO algorithm and Box-Jenkins time series

Industry	Average of actual stock price	Average of stock price prediction with PSO algorithm	Error percentage	Average of stock price prediction with Box-Jenkins time series	Error percentage
Cement	7982	8003	0.006	8132	0.021
Textile	4363	4359	0.006	4459	0.018
Transportation	4977	4986	0.002	4855	0.017
Sugar	3696	3692	0.004	3490	0.027
Fireproof products	6000	5996	0.003	4726	0.13
Chemical industries	8197	8176	0.002	8345	0.009
Mines and metals	1732	1745	0.002	1520	0.017
Medicine	1558	1554	0.005	1105	0.19
Food products	3466	3394	0.007	3459	0.016
Others	4979	4928	0.008	4925	0.005

5. Discussion and Conclusions

The purpose of this research is to predict stock prices using Particle Swarm Optimization Algorithm and Box-Jenkins time series. Research population includes 165 corporations in 16 years-period (2001-2016). The relevant regression equation was created using two variables of earnings per share and price to earnings per share, and stock prices were predicted through particle swarm optimization algorithm in MATLAB. IBM SPSS was used to predict stock prices with Box-Jenkins time series. Results indicate that particle swarm optimization algorithm with 4% error and Box-Jenkins time series with 19% error; have the potential to predict stock prices of companies. Moreover, PSO algorithm model predict stock prices more precisely than Box-Jenkins time series.

It is recommended to investors, financial institutions, finance suppliers, loaners and those who stock price prediction is important for them, to use particle swarm optimization algorithm for prediction of stock prices. Also they should be cautious about internal and external factors when implementing this algorithm. In addition, it is recommended to researchers and financial statements analyses to use PSO algorithm as prediction tools in other filed of accounting and auditing. Lack of corporate data in several years is the main barrier in implementation of this model. As such, among many effective factors, we only chose earnings per share and price to earnings per share.

References

- 1) Alborzi, M., Yaghoob Nejad, A., Maghsoudi, H. (2008). Application of Artificial Neural Network to predict the stock price index and cash returns. *Emp Stud Finan Accou*, 22, 119-127.
- 2) Jahankhani, A., Safarian, A. (2003). Stock market reaction to the estimated dividend per share in the Tehran Stock Exchange. *J Financ Res*, 5(2), 7-23.
- 3) Kenedy, J., Eberhart, R. (1995). Particle swarm optimization in proc. IEEE International Conference on Neural Networks, 4, 1942-1948.
- 4) Karegowda, G., Kumari, S. (2013). Particle swarm optimization algorithm based k-means and fuzzy c-means clustering. *Int J Adv Res in Com Sci & Soft Eng*, 13(7), 448-451.
- 5) Lin, W., Lian, Z., Gu, X., Jiao, B. (2014). A local and global search combined particle swarm optimization algorithm and its convergence analysis. *Math Pro in Eng* 2014, Article ID 905712, <http://dx.doi.org/10.1155/2014/905712>.
- 6) Mehrani, S., Mehrani, K. (2003). The relationship between profitability ratios and stock returns in the Tehran Stock Exchange. *Iranian Account & Audit Rev*, 10(3), 93-105.
- 7) Nobel, J. (2008). Europe sugar policy crisis Agricultural, Industrial and Economics Beetroot and Cane of Iranian Sugar Industry, 15, 1-30.
- 8) Oliveira, F., Nobre, C., Zarate, L. (2013). Applying artificial neural networks to prediction of stock price and improvement of the directional prediction index—case study of PETR4 Petrobras Brazil. *Ext Syss with Appl*, 40 (18), 7596–7606.
- 9) Sharifi, E., Verij Kazemi, M., Tamandani, S. (2011). Improved detection capacity distributed generation units to improve power quality indices using the improved PSO algorithm. *Iranian J Marine Sci & Tech*, 56(56), 67-75.