

Forecasting Initial Public Offering Pricing Using Particle Swarm Optimization (PSO) Algorithm and Support Vector Machine (SVM) In Iran

Shaho Heidari Gandoman (Corresponding author)

Department of Accounting, Sanandaj Branch, Islamic Azad University, Sanandaj, Iran

E-mail: shahoheidari@yahoo.com

Navab Kiamehr

Department of Accounting, Boroujerd Branch, Islamic Azad University, Boroujerd, Iran

Mahmood Hemetfar

Educator, Department of Accounting, Islamic Azad University, Boroujerd, Iran

Received: November 4, 2016 Accepted: November 30, 2016

doi:10.5296/ber.v7i1.10910 URL: <https://doi.org/10.5296/ber.v7i1.10910>

Abstract

The present study compares the ability of neural networks, support vector machine, and model derived from combining particles swarm optimization (PSO) algorithm and support vector machine (SVM) to forecast the initial public offering pricing. The purpose of this research is to design a model that helps investors recognize the validity of the initial public offering pricing and hunt profitable opportunities. The variables used in this study are selected among those variables which are in the disposal of investors who have limited access to information before the offering. On the other hand, these results can be useful for publishing companies, admissions consultant, underwriting and legislators of the stock exchange. We have considered the ninth day offering prices, since volatilities are gone and prices seem to be more realistic. The results show that the combination of particle swarm optimization (PSO) algorithm and support vector machine (SVM) markedly increases the forecasting power. As a result, support vector machine models can increase the accuracy of initial public offering pricing and provide significant economic benefits as reducing less than real pricing costs.

Keywords: Business, Finance, Initial public offer, Pricing, Neural networks, Support vector machine, Particle swarm optimization algorithm

JEL Codes: C81, D44, G12, H80

1. Introduction

Support Vector Machine is widely used in the formulation of non-linear relationships in a wide range of decision-making issues. Due to the special capabilities of support vector machine, it is expected to be used in complex issue of initial public offerings (IPO) or in a sense, the first offer of company shares to the public through the stock exchange, since the pricing in initial public stock offerings involves a very difficult and complex decision process that is done on the basis of inadequate information.

In addition, the valuation of a company in order to determine the initial public offering price depends on many variables which are independent of each other and the relationship between them is not clear. In such circumstances, the pricing of initial public offerings is difficult for the enterprise and the supplier company. The fact that many of the initial public offering prices are less than real price the fact is proven (Ross, 2003). Although there are many reasons for the occurrence of such a phenomenon, the performance of initial public offerings pricing, especially for the supplier company is still desirable and important. One of the main issues in the pricing of initial public offerings is an assessment of the company.

To assess the value of the firm, the value of tangible and intangible assets of the company should be calculated. Intangible assets such as the future growth of the company or the industry in the near future are difficult to forecast. In fact, the accuracy of the initial public offering price will be determined only after the offering in the market (Lubeck, 2001).

In recent years, support vector machine technology is widely used in the areas of Finance and Accounting to improve the quality of decision making. Application of support vector machine is mostly in the scope of issues that involve forecasts based on pattern matching or classification. One of the distinctive characteristics of the support vector machine is its ability to provide a reasonable response, even assuming the data is incomplete or abnormal (Reber, 2005). Since financial data are often incomplete or abnormal, support vector machine technology becomes more important.

The present study compares the forecasting power of support vector machine, and the model derived from the combination of the particle swarm optimization algorithm and support vector machine to forecast the initial public offering pricing. The purpose of this research is to design a model that helps investors recognize the validity of the initial public offering pricing and hunt profitable opportunities. The variables used in this study are selected among those variables which are in the disposal of investors who have limited access to information before the offering. On the other hand, these results can be useful for publishing companies, admissions consultant, underwriting and legislators of the stock exchange.

2. Problem Statement

During the initial public offer of shares to the market, there is little knowledge about them. So

the initial public offering market is at risk of adverse selection. On the other hand, suppliers believe that underpricing is necessary in order to encourage investors to buy stocks and reduce marketing costs (Norsys, 2003). One of the most common theoretical explanations for the problem of underpricing is signal-based models.

Signaling theory, introduces price as a tool for signaling. Individuals within the organization may try to transfer information on company's value to external investors. The insufficient information and signals that may be interpreted indicates that there may probably be good opportunity to buy profitable stocks (Menhaj, 2007).

This study examines the capability of support vector machine model to provide better forecasts for the rate of return on the first day of initial public offerings. Grounded theory used in the research is asymmetric knowledge of economic activists; that means that a number of investors have greater awareness about the true value of the shares sold in the initial public offering, while others remain unaware since it is more difficult and costly for them to acquire further information. The purpose of the current study is to examine whether forecast power of the model can be increased by entering assumptions of this theory into support vector machine.

Support vector machine is able to extract relevant information to create meaningful relationships and thus provide better forecasts compared to agreement economic models. Support vector machine, unlike linear models, reflects nonlinear effects and complex interactions between signals of company's value (Reber, 2005).

2.1 The Importance of this Research

The initial public pricing issue is complex and sensitive. The accuracy of the accurate market of initial public offerings would be clear just after the initial public offering of the company in the market (Ross, 2003). Many companies incur huge losses due to the inability to determine the correct pricing of initial public offerings. Although all initial public offerings are not experiencing this dilemma, to have a tool to maximize the forecasting power of underpricing is strongly felt.

From the perspective of investors, correct market value forecast is certainly a valuable tool. Therefore, forecasting initial returns is important. In Iran, for pursuing policies of Article 44 and the transfer of state-owned enterprises and the specific issues in the pricing of these companies, the importance of further investigation is even more.

However, because of the nonlinear relationships and complex interactions of firm value, investing in initial public offerings stocks requires the ability to forecast higher returns than existing models. Support Vector Machine is able to establish meaningful information from the data and so provide better forecasts compared to economic models. Unlike linear models, reflect nonlinear effects and complex interactions between signals of company's value (Reber, 2005).

2.2 Research Objectives

The main goal of this research is to create models to forecast the pricing of initial public

offerings and to compare their explanatory power. These models are created using artificial neural networks, support vector machine, and particle swarm optimization algorithms. So it is tried to provide a useful approach to investing in initial public offerings.

2.3 The Statistical Population

The population of this research includes all companies listed in Tehran Stock Exchange for a period of 5 years from 1390 (2011-2012) to 1394 (2005-2016) based on the information in the Rahavard Novin and Omidnameh softwares. Due to lack of access to detailed financial information on companies out of stock, the population of this research includes companies listed in Tehran Stock Exchange during the period 1390 to 1394.

2.4 Research Hypotheses

The first hypothesis - Forecast strength of support vector machine (SVM) is higher than the neural network to forecast the rate of return on the first initial public offering.

The second hypothesis - Using particle swarm optimization (PSO) algorithm to determine the appropriate parameters, increases the performance of support vector machine (SVM) model to forecast the rate of return on the first day of initial public offerings.

Table 1. Research variables

Collection method	Description	variables
Rahavard Novin	Pricing	price
Omidnameh	Average Earnings per share (before tax) over three fiscal period before offering	LEPS
Omidnameh	Current ratio before offering	curnt
Omidnameh	Quick ratio before offering	Quik
Omidnameh	Reverse sales during the financial period before offering	Sale
Omidnameh	Return of assets before offering	Roa
Omidnameh	Return of equity before offering	Roe
Omidnameh	Net profit margin before offering	npm
Omidnameh	Operational profit margin before offering	opm
Omidnameh	Asset turnover before offering	Atu
Omidnameh	(Non) disclosure of the anticipated dividend displayed with variable zero – one so that if such disclosure is made we use one, otherwise zero is used.	DIS
Omidnameh	Expected profit per share for the next year I omidnameh	EPS
Omidnameh	Dividend through interest of previous year in Omidnameh	DIV
Rahavard Novin	Expected return: calculated based on CAPM that reflects B of the industry	ER
Rahavard Novin	The volatility of the industry: Using the coefficient of variation of daily returns in various industries over the period of 30 trading days prior to the release calculation.	V

3. Research Methodology

3.1 Descriptive Statistics

Data analysis to verify the research hypotheses and research questions are of particular

importance. Today, most research is based on data collected on the subject, and data analysis of is the main and most important part of the research. The raw data are analyzed using appropriate software and after processing, are offered to the user in the form of information. In analyzing the collected data, descriptive statistics to examine and describe the variables are first expressed.

Table 2. Descriptive statistics

Max.	Min.	Standard deviation	Deviation from the average	average	Index variable
23544	1025	5430.599	975.364	4634.35	Price
13921	67	2487.791	446.821	1097.61	Leps
0.000010	0.0000000010	0.00000179	0.000000032	0.0000006	Sale
2.98	0.13	0.6497	0.116	1.153	Current
2.65	0.07	0.6289	0.1129	0.9842	Quick
3.69	0.001	0.6688	0.120	0.2499	Roa
4.020	0.014	0.8449	0.1517	0.4648	Roe
0.930	0.004	0.2550	0.04580	0.2517	Npm
3.000	0.004	0.5567	0.09999	0.3853	Opm
9.300	0.005	1.751	0.3145	0.8932	Aut
12892	98	2277.711	409.089	934.26	Eps
10096	12	1812.306	325.500	724.94	Div
13.88	-0.15	2.5814	0.4636	2.8352	Er
0.800	0.001	0.2453	0.044	0.2491	V

This table shows values obtained for central indices and dispersion of the variables.

To test normality of data, Kolmogorov - Smirnov test is used and the results are as follows.

H0: the assumption of data normality is confirmed.

H1: the assumption of data normality is not confirmed.

Table 3. Test normality of data

p-value	Kolmogorov – Smirnov statistic	Standard deviation	average	variable
0.038	1.409	5430.599	4634.35	Price
0.002	1.890	2487.791	1097.61	Leps
0.000	2.057	0.00000018	0.0000006001	Sale
0.397	0.897	0.64970	1.1535	Current
0.358	0.926	0.62893	0.9842	Quick
0.000	2.192	0.6688	0.24994	Roa
0.001	1.899	0.844924	0.46487	Roe
0.047	1.368	0.255037	0.25177	Npm
0.046	1.374	0.556735	0.38532	Opm
0.001	1.993	1.7516	0.89323	Aut
0.001	1.986	2277.711	934.26	Eps

0.001	1.932	1812.306	724.94	Div
0.163	1.120	2.58145	2.8352	Er
0.157	1.128	0.2453	0.2491	V

According to the values obtained for p-value, where the significance level (p-value) is greater than 0.05, the null hypothesis is confirmed, and indicates that the data follow a normal distribution. In these cases the variables current, quick, ER and V are normally distributed.

Normality test results are shown in the following table.

Table 4. Normality test results

p-value	Kolmogorov – Smirnov statistic	Standard deviation	average	variable
0.684	0.716	0.85066	8.03	LnPrice
0.954	0.514	1.2367	6.0812	LnLeps
0.699	0.707	2.180	-16.242	LnSale
0.788	0.653	1.6645	-2.7044	LnRoa
0.895	0.576	1.1452	-1.4969	LnRoe
0.215	1.055	1.592	-2.1686	LnNpm
0.130	1.170	1.6972	-1.8889	LnOpm
0.691	0.712	1.5135	-1.0860	LnAut
0.332	0.947	1.0425	6.0346	LnEps
0.951	0.518	1.42151	5.4655	LnDiv

Normality test results showed a significant level more than 0.05, where data follow a normal distribution (normality assumption is confirmed).

To test the correlation of variables, Pearson correlation test is used. The results showed that the variable "price" is correlated with variables er, leps, roa, roe, eps. The results are shown in the table below.

Table 5. Pearson correlation test results

p-value	The Pearson correlation coefficient of variable price with other variables	variable
0.000	0.721	Leps
0.537	0.115	Sale
0.069	-0.331	Current
0.204	-0.235	Quick
0.000	0.625	Roa
0.004	0.506	Roe
0.326	0.182	Npm
0.523	0.119	Opm
0.081	0.318	Aut
0.000	0.763	Eps
0.082	0.318	Div
0.019	0.418	Er

0.101	-0.300	V
-------	--------	---

The correlation coefficient between variable price and variable eps and leps is 0.763 and 0.721, respectively, which are highly correlated with each other.

The first hypothesis - Forecast strength of support vector machine (SVM) is higher than the neural network to forecast the rate of return on the first initial public offering.

To test the first hypothesis, aimed to compare the forecasting power of the neural networks and support vector machines, first, we forecast prices based on neural networks, and then for the same data we use support vector machine.

3.2 Neural Network Model

Neural network model considered for this research is a multilayer neural network. According to the experiments that have been conducted, the best structure for the network is a two-layer neural network where the intermediate layer has 7 neurons. The transfer function of the middle layer is Sigmund function and for the output layer is a linear function. The learning algorithm of back-propagation is used to train the network with 0.15 learning rate. As also shown in the figure below, the training is stopped after 7 cycles, and the cessation reason is clear based on the data validation curve, testing and training.

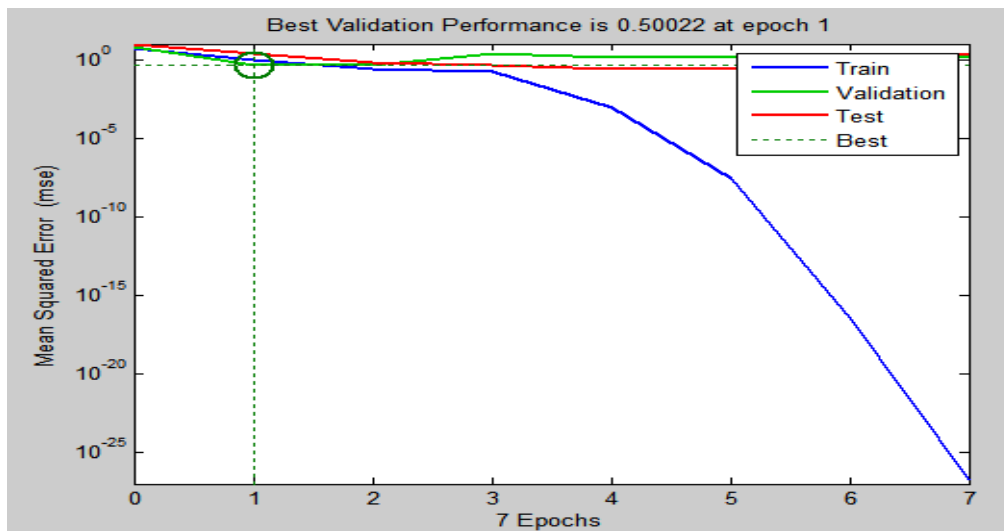


Figure 1. Neural network model result

In the following curve, explanatory power for neural network training data set is shown. In this chart, the x-axis is actual output and y-axis is the desired output of the network. The dispersion of the points around the curve $y = x$ is less, the forecasting power of the neural network is more.

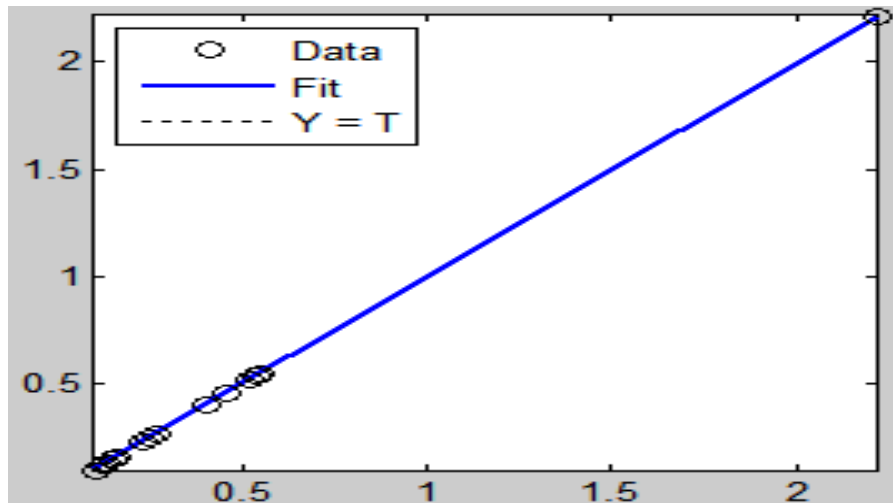


Figure 2. Explanatory power for neural network training data

In the figure below the forecasted price based on support vector machine is shown.

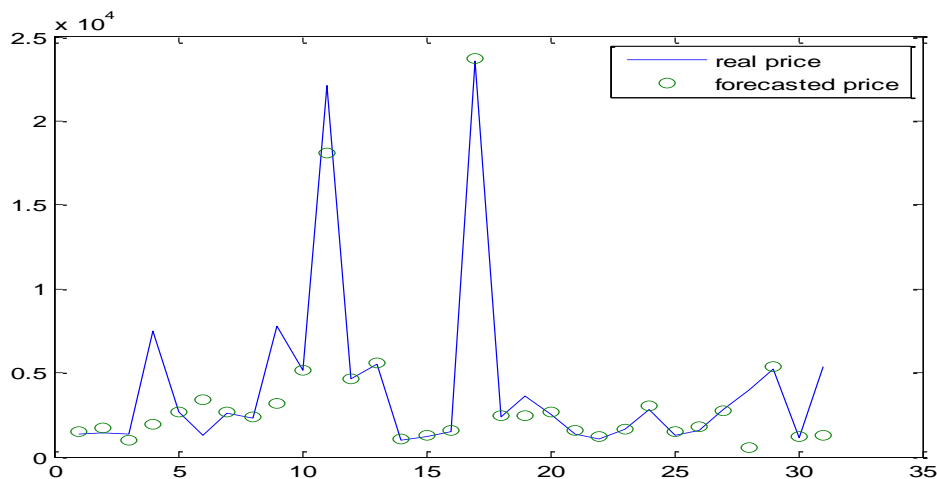


Figure 3. Forecasted price based on support vector machine

3.3 Support Vector Machine Forecast Model

Support vector machine (SVM) is one of the most powerful and innovative algorithms in the field of data classification. This algorithm is built based on statistical learning theory and its purpose is to generalize new samples, what is a weakness in techniques such as artificial neural networks. The figure below shows the structure of a support vector machine.

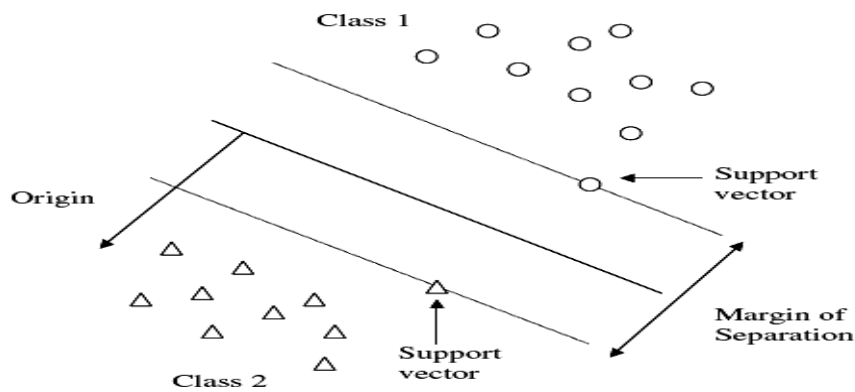


Figure 4. The structure of a support vector machine

SVM is used to separate the linear and nonlinear separable data. In the separation of inseparable non-linear data, support vector machine transfers data into higher dimensional space, so that data in that space linearly have the highest resolution. This is done based on kernel functions as unique advantages of support vector machine. Another important feature of SVM is to find the maximum distance between the parallel plates compared with each other, so that the minimum number of training samples be placed between the two plates. This distance between the plates is called separation margin. The combination of parallel plates and their margins is called decision boundary.

The main objective of the training phase of SVM is solving an optimization problem to maximize the margin while minimizing the number of training samples located within the margins. If using kernel functions, the decision function is as Equation (1):

$$g(x) = \text{sign} \left(\sum_{i=1}^{N_s} \alpha_i y_i k(X_i, X) + b \right) \quad (1)$$

In relation (1), $X_i \in R^d, i = 1, \dots, N$, d is the dimension of features, N is the number of training samples in two categories, $y_i \in \{-1, +1\}$, α_i are the coefficients obtained from solving the optimization problem, and N_s is the number of training samples containing α_i coefficients of zero, and are exactly located on plates. These samples are called support vectors (SV) and are important training data because the SV alone would suffice to define the margins of resolution.

In this study, the Gaussian RBF kernel, is applied as the kernel SVM function according to the data shown good results in classification. The main problem is to set this kernel parameter to determine the best separation zone between two categories.

In the figure below, the forecast price based on support vector machine is shown.

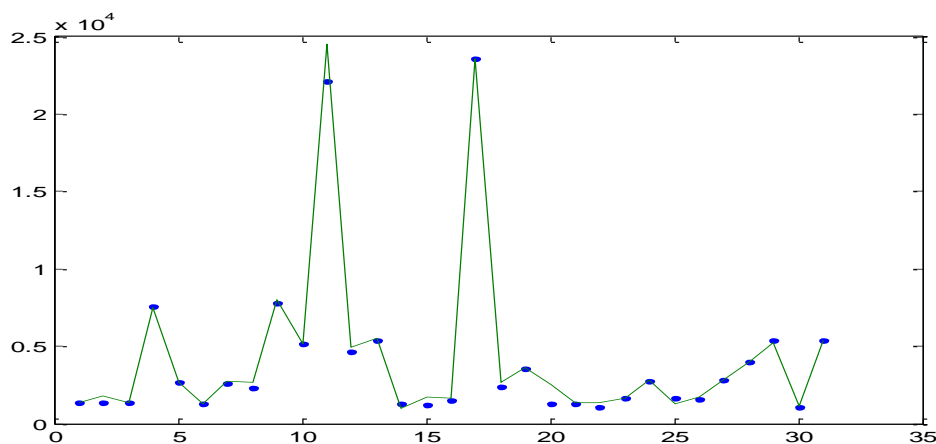


Figure 5. The forecast price based on support vector machine

The performance indicators for neural network model (MLP) and support vector machine (svm) are shown in the following figure. For each set of data, three functional indices are reported including mean absolute error (MAE), root mean square error (RMSE) and the sum of squared errors (SSE).

Table 6. The performance indicators for neural network model (MLP) and support vector machine (svm)

SSE	RMSE	MAE	model
0.7321	0.09980	0.07015	MLP
0.7590	0.1000	0.07123	SVM

It can be seen that based on performance indicators considered for two models, MLP for all three indicators show better performance compared to SVM. Therefore, we can conclude that the MLP has a higher explanatory power than the SVM. Thus, the first hypothesis is rejected.

The second hypothesis - Using particle swarm optimization (PSO) algorithm to determine the appropriate parameters, increases the performance of support vector machine (SVM) model to forecast the rate of return on the first day of initial public offerings.

In this stage of the study, in order to test the second hypothesis, the particle swarm optimization algorithm is used to select the parameters c and gamma suitable for support vector machine.

3.4 Optimization of Support Vector Machine Using Particle Swarm Optimization Algorithm

One of the biggest problems in the use of intelligent methods such as support vector machine is that parameters of these systems should be configured manually using trial and error. Considering that these parameters are depending on the case, and the parameters that are suitable for case A may not be suitable for case B, in this study, using a smart method, the best parameters are chosen. This method combines the particle swarm optimization algorithm and support vector machine. This is the first time our method is used to forecast prices

(according to the review that we have done), works in the way that in each iteration, parameters are offered from PSO to SVM and SVM by using them forecast price, and in the next iteration, PSO its specific logic, offers the new values for the parameters and this approach continues as long as the error rate reaches to a minimum. We continue this procedure until the desired minimum error is achieved, or after a specific number of repeats, the value of error is not greatly improved, and continuing the repeats does not have logical usefulness.

In this paper we forecast prices based on RBFSVM that has two parameters of Gamma and C. [AlphaY, SVs, Bias, Parameters, nSV, nLabel] = RbfSVC (Samples, Labels, Gamma, C)

In normal mode, the values should be determined manually using trial and error. In this study, we determined the values of these parameters by using PSO. As can be seen in the results of the simulation, the forecast results are significantly improved. The following figure shows considered a model for forecasting schematically.

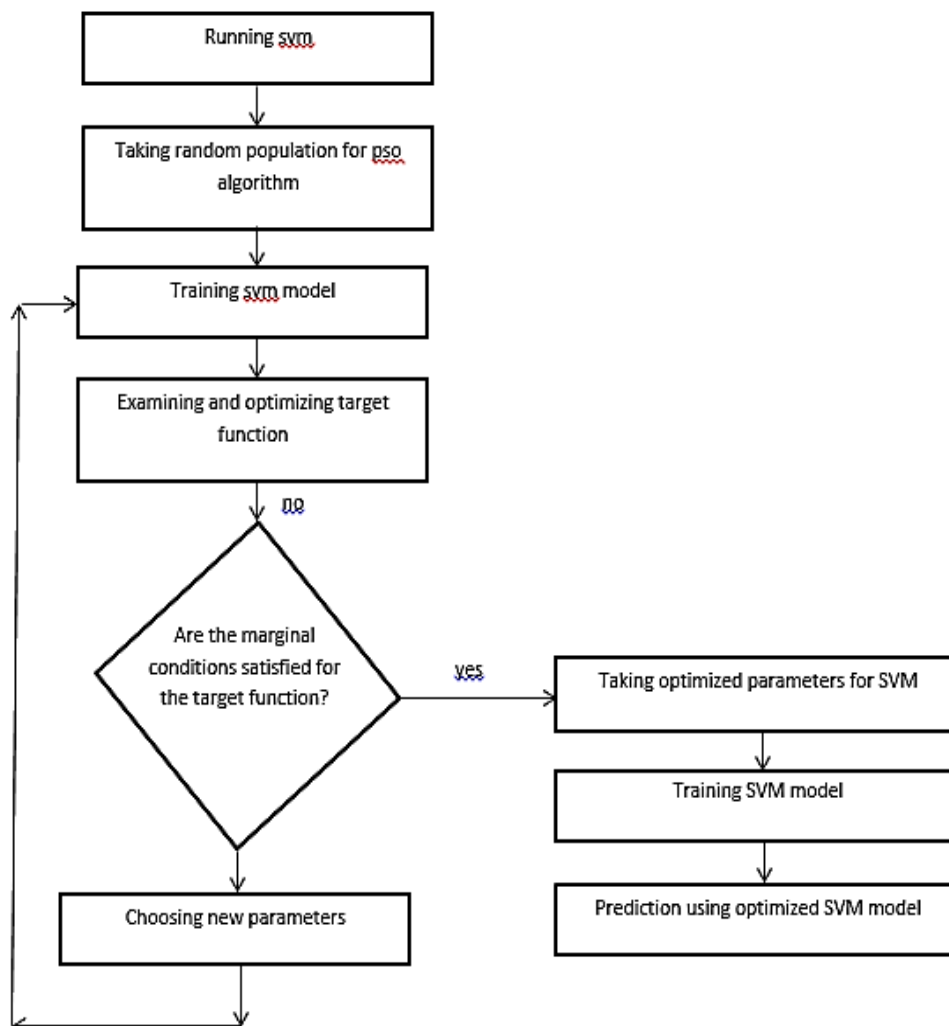


Figure 6. Schematic of the proposed SVM-PSO

To study the effect of a number of iterations (epoch) we first run the program with one repeat,

and write the results. Each time, we increase the number of repetitions once, and compare the results with the previous stage. Comparing the results indicates that increasing the number of iterations corrects the forecasted outcomes. The effect of increasing the number of iterations is converged after 25 iterations and more repeats do not have a significant effect on the results of the forecast.

The following figures present optimization parameters and compare the forecasted results with the actual values using SVM-PSO technique.

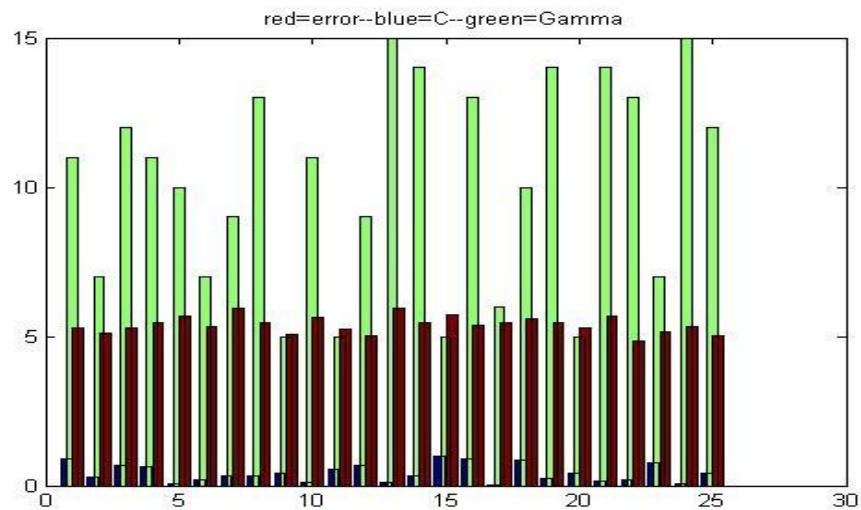


Figure 7. optimization parameters and compare the forecasted results with the actual values using SVM-PSO technique

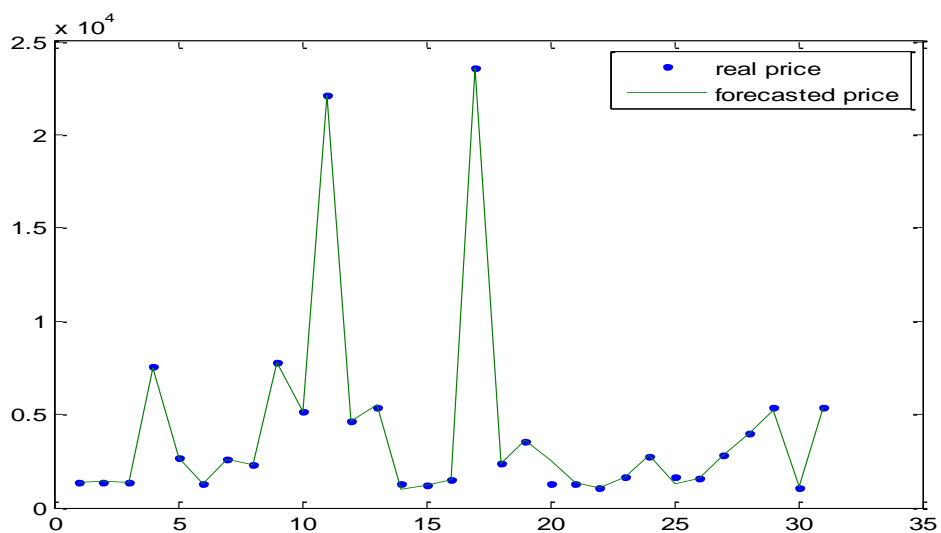


Figure 8. Optimization of parameters of support vector machine and related error

Table 7. The results of MLP, SVM and SVM-PSO models

SSE	RMSE	MAE	model
0.7321	0.09980	0.07015	MLP
0.7590	0.1000	0.07123	SVM
0.7358	0.0856	0.06235	SVM-PSO

Given the results obtained in the table above, it can be seen that SVM-PSO method works better than the SVM and MLP models, and the forecast based on this model has less error. Thus, the second hypothesis is confirmed.

4. Conclusion

The overall objective of this study is to compare the forecast power of neural network models and support vector machine (SVM) in order to forecast the output of the first day of initial public offerings and to provide a model of particle swarm optimization (PSO) algorithm in order to help and train the support vector machine to forecast the best price. The idea that markets are affected by information asymmetry and as a result cares about signals such disclosures of companies in their Omidnameh, forms the theoretical framework of this study.

This study aims to establish a proper forecast tool to forecast the pricing of initial public offerings by a particle swarm optimization (PSO) algorithm and support vector machine (SVM). The results indicate that the particle swarm optimization (PSO) algorithm and support vector machine (SVM) for selecting optimum variables that leads to the selection of independent variables, average earnings per share during the financial period before offering (leps), expected profit per share for the next year inomidnameh (eps), reverse sales during the financial period before offering (sale), expected returns (er) return of equity before offering (roe), and operational profit margin before offering (opm), significantly increases forecast power compared to the conditions where support vector machines is used alone.

Thus we come to the conclusion that the combined particle swarm optimization (PSO) algorithm and support vector machine (SVM) in order to select the optimal variables, leads to a dramatically better estimation of the expected pricing. Particle Swarm Optimization (PSO) algorithm increases the accuracy of initial public offerings pricing and thus provides significant economic benefits as reduces underpricing costs. Artificial neural networks are known as black boxes testing, despite their ability to detect relationships between variables, do not indicate to the user how these relations are. Thus the type of relationship between the dependent variables and the independent variables (in terms of intensity and direction of the relationship) is not clear. However, this model was used to compare support vector machine model comparison and better performance of this model.

References

- Lubic, H. Y. (2001). Initial public offering prediction using neural network. Doctoral dissertation, George Washington University.
- Minhaj, M. B. (2007). Principles of Neural Network (computing intelligence), Tehran University of Technology (Tehran Polytechnic) Fourth Edition

Nvrsys, M. (2003) Analyzing data in SPSS110 education, translations of Ftohi.A &. A.F. Kanvn, A.F. Teheran Science Publishing.

Reber, B., Berry, B., & Toms, T. (2005). Predicting mispricing of initial public offerings. *Intelligent Systems in Accounting, Finance and Management*, 13, 41-59.

<https://doi.org/10.1002/isaf.253>

Ross, S. A., Westerfield, R. W., & Jordan, B. D. (2003). Fundamentals of corporate finance. McGraw-Hill/Irwin, 6.

Copyright Disclaimer

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).