A New Approach to Forecasting Stock Price with EKF Data Fusion

H. Haleh, B. Akbari Moghaddam, and S. Ebrahimijam

Abstract— **Obtaining to the method with the least prediction error is one of the challenging issues of financial and investment markets analyzers. Investors often use two different views of technical and fundamental analysis of prices for buying and selling their desired shares. But each of these two methods alone may have not enough performance due to differences between the actual value of the share and its market price.**

This paper presents a predictive model named extended Kalman filter which simultaneously fuses information and parameters of technical and fundamental analysis. Then as a real test, the model implemented for the shares of one of industrial company in Iran. Finally, the obtained results will be compared with other methods results such as regression and neural networks which shows its desirability in short-term predictions

Index Terms—**Stock exchange, data fusion, Extended Kalman filter, technical and fundamental analysis.**

I. Introduction

The development and economic growth in each country has direct relation with the investments done in that country. So that in the developing country, capital formation is considered as an important factor in economic growth and making the process for development.

In developed countries stock exchange is an official organization for capital, through attracting investment and stagnant savings of the society and conduct their affairs in production which led to increase economic growth and prosperity has been produced and this way effect on macroeconomic variables such as GDP, money, inflation and interest rates. That's why stock market boom and mobility known as a criterion of dynamic economies of the countries.

Today's, investing in stock is an important part of society and economy, in the other hand, stock price volatility in the all stock exchanges is commonplace. Stock prices affected by external and internal factors like political, economic and etc, so the stock price prediction for capitalists is very important to be able to return most of their investment will earn.

Meanwhile, given the unavailability of accurate information about the factors affecting stock market fluctuations and of unknown factors affecting stock price changes, due on bringing people to predict price changes by the companies but this simply not possible.

Generally there are two viewpoints of analyzing: Technical analysts and fundamental analysts [1]. Technical analysts can calculate the intrinsic value of stock. They believe that the market has undergone a pseudo psychological mode and history is always repeatable and patterns at any time cause reputation of trends of price. Therefore by studying the past trend, the future could be predicted [2]. But in terms of fundamental analysts, the stock market has no memory and prices are changed randomly.

As noted, the main purpose of technical analysis is predicting trends of stock price. However, predictions are often not correct and have some errors that the rate decreases with increasing of information. The other comment which was raised said: technical analysis based on the weak principles. For example, the expectation that some of historical patterns of the share prices will be repeated in the future may not necessarily occurs, because market conditions will change over time.

But in 1970, Fama presented a new hypothesis which is called efficient market hypothesis (EMH), it is commonly believed that in each period of time, asset prices such as stock prices reflect all information about them [3]. And all Technical or fundamental analysis and prediction are failed. But this simple idea was a lot discussed of controversial financial investigators.

In many of the arguments raised against the theory of market efficiency, there have been identified and repetitive patterns for prices, there are examples of this case, we can mention like: January effect or Blue Monday in Wall Street which are some of those predictable repetitive patterns [4].

According to a research done in Iran stock exchange market, very poor efficiency was dominated and a specific pattern of behavior can be observed in the changing of prices, so awareness of this model can help investors to obtain more benefits [5].

In the recent years many researches were done in this area, some tried to resolve the problem by classical methods such as exponential smoothing, trend analysis, and regression equations [6], some uses intelligent methods such as neural networks and fuzzy or combination methods [7]. But a lot of bugs to any of these methods are entered. as a further example linear prediction models are used in the methods, whereas in reality most processes are nonlinear. Or intelligent methods require a lot more data for the learning system. in the absence of data; models of changes are not correctly identified and systems shall operate incorrectly in non-trained conditions. Or even the difference between the

H. Haleh is with Faculty of Industrial and Mechanical Engineering; Islamic Azad University-Qazvin Branch, Qazvin, Iran; e-mail: hhaleh@cc.iut.ac.ir

B. Akbari Moghaddam, is with Faculty of Management and Accounting Science, Islamic Azad University-Qazvin Branch ,Qazvin, Iran; e-mail: finan@qiau.ac.ir.

S. Ebrahimijam is with Faculty of Industrial and Mechanical Engineering; Islamic Azad University-Qazvin Branch , Qazvin, Iran; e-mail: ebrahimijam@mrl.ir

expected price stock market prices and lack of adherence to the theoretical model, prevent correct prediction of prices.

This study presented a model, based on technical analysis in stock market prices. Method used in this study is kind of time series entitled the Kalman filter which acts based on analyzing change of data versus time. But according to the past researches on stock market areas, the use of linear models are not capable of predicting prices in the long time, so a kind of advanced methods which is called extended Kalman filter algorithm will be used, The advantage of developing this type of filter is able to use non-linear models as main system model [8]. By the data fusion method of extended Kalman filter the possibility of combining different indicators used in the fundamental analysis like economic variables with technical analysis and general trend of change is possible.

II. Extended Kalman Filter

Assume that we are going to accurately measure a parameter, like moving distance of the car in the road. There is a physical-mathematical model for moving with constant acceleration which is $x = \frac{1}{2}at^2 + V_0 t$. Then, Are you sure the obtained answer from the model is the real distance? Obviously no!! You didn't take accounts the deviation from the straight road because of the holes, frictions and etc. So, you need to measure that with measuring devices like odometers (z_1) or GPS (z_2). But, which one is correct and accurate (z_1 or z_2)? The answer is maybe both, but each one with its error tolerance. So, it is needed to explore the accurate answer by fusing the results. One of the well-known fusion methods is called Kalman filter.

Kalman filter is created from the name Rudolf E. Kalman in an article which was published in 1960 that presents recursive solution to filter the linear discrete data [9].

Kalman filter actually is a set of mathematical equations that is type of optimally estimator, predictor and corrector which sensibly minimizes the estimation error covariance [10].

Kalman filter is suitable for the data which are distributed normally according to below equation.

$$X = N(\mu, \delta^2) \tag{1}$$

$$\mathrm{f}_{\mathrm{x}}(x) = \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\delta^2}} \tag{2}$$

The computational origin of the Kalman filter is mentioned in the next equations. This is the simple way of two parameter fusion.

$$\mu = [\sigma_{z_2}^2 / (\sigma_{z_1}^2 + \sigma_{z_2}^2)] z_1 + [\sigma_{z_1}^2 / (\sigma_{z_1}^2 + \sigma_{z_2}^2)] z_2 \tag{3}$$

$$\hat{X}(t_2) = \mu \tag{4}$$

Where, μ is weighted average of two measuring devices and $\sigma_{z_n}^2$ is the variance of the error in measurement for each device and $\hat{X}(t_2)$ is estimated value of the measuring parameter in t_2 time.

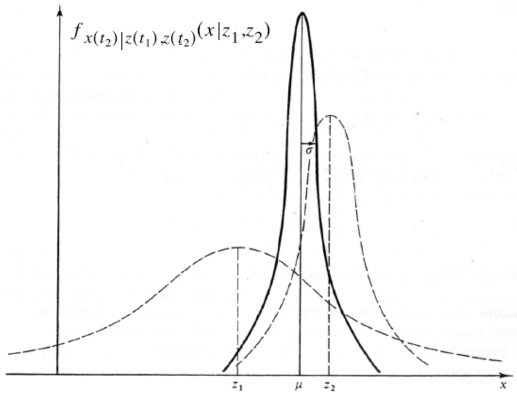

Figure 1. Conditional density of position based on data z1 and z2 (S.Maybeck [17])

So,

$$\begin{aligned} \hat{x}(t_2) &= [\sigma_{z_2}^2 / (\sigma_{z_1}^2 + \sigma_{z_2}^2)] z_1 + [\sigma_{z_1}^2 / (\sigma_{z_1}^2 + \sigma_{z_2}^2)] z_2 \\ &= z_1 + [\sigma_{z_1}^2 / (\sigma_{z_1}^2 + \sigma_{z_2}^2)][z_2 - z_1] \end{aligned} \tag{5}$$

The following equation is explored from the top equation.

$$\hat{x}(t_2) = \hat{x}(t_1) + K(t_2)[z_2 - \hat{x}(t_1)] \tag{6}$$

And the Kalman coefficient (K) is calculated from below formula.

$$K(t_2) = \sigma_{z_1}^2 / (\sigma_{z_1}^2 + \sigma_{z_2}^2) \tag{7}$$

$$1/\sigma^2 = (1/\sigma_{z_1}^2) + (1/\sigma_{z_2}^2) \tag{8}$$

The term which minimizes the estimation error covariance is:

$$\sigma_x^2(t_2) = \sigma_x^2(t_1) - K(t_2)\sigma_x^2(t_1) \tag{9}$$

Kalman filter is trying to understand the general issues which are to estimate the state of $X \in R^n$ discrete-time control process that has linear differential function.

Kalman filter can be applied for linear systems, but linear system doesn't exist. "…Now what happens if a process that should be estimated and (or) relationship with the measurement process is nonlinear? Many successful and interesting applications in the Kalman Filter are in these conditions." One method of nonlinear kalman filtering is to expand the nonlinear terms of the system equation in a Taylor series expansion around a nominal point of $\bar{x}$. The other method is using linearized Kalman filter which starts with a nonlinear system, and then find a linear system whose states represent the deviations from a nominal trajectory of the nonlinear system. Here, the first-order Taylor series used for expanding the state equation and output equation around a nominal state. So we have to know the nominal trajectory of the states $\bar{x}$ [11].

"…The Kalman filter that linearizes about the mean and covariance of states is known extended Kalman filter (EKF)" [12].

"…The idea of the EKF (Extended Kalman Filter) is to use the stimulation of x as the nominal trajectory in the linearized Kalman filter" [11].

Assume that the process has $X \in R^n$ state vector. And the process is given with a nonlinear random differential equation.

$$x_k = f(x_{k-1}, u_{k-1}, w_{k-1}) \tag{10}$$

By measurement $z \in R^m$ in which:

$$z_k = h(x_k, v_k) \tag{11}$$

Where the random variable w_k and v_k represents the process and measurement noise. In this equation, nonlinear function of f in the differential equation 1 related to the state in the previous time step to current time k .These are also include function parameters u_k and zero mean noise process (w_k). Nonlinear function h in measurement equation 11 is related to the state x_k to the measurement z_k [13].

There are some parameters like, A, H and C.

A is the Jacobian matrix of partial derivatives of f with respect to x .

$$A_{[i,j]} = \frac{\partial f_{[i]}}{\partial x_{[j]}}(\hat{x}_{k-1}, u_k, 0) \tag{12}$$

C is the Jacobian matrix of partial derivatives of f with respect to x .

$$C_{[i,j]} = \frac{\partial h_{[i]}}{\partial x_{[j]}}(\tilde{x}_k, 0) \tag{13}$$

Also, there are some variables like P which needs to set initial value, to initiate this variable it is needed to define a priori and a posteriori estimate errors as

$$\tilde{e}_{x_k} = x_k - \tilde{x}_k \tag{14}$$

$$\tilde{e}_{x_k} = z_k - \tilde{z}_k \tag{15}$$

The initial value for The a priori estimate error covariance is then

$$P_k^- = E[e_k^- e_k^{-T}] \tag{16}$$

The execution equations of Kalman filter are mentioned as below.

$$K_k = P^- C_k^T (C_k P_K^- C_k^T + R)^{-1} \tag{17}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - h(\hat{x}_k^-, 0)) \tag{18}$$

$$P_k = A_k(I - K_k C_k H) P_k^- A^T{}_k + Q \tag{19}$$

III. Data Fusion Model By EKF

According to the previous Sections about the trend of change in stock price and the reasons offered by the stock market technical analysts, it seems that there is a strong relationship between price changes and the time. But according to fundamental analyst's comments, the internal and external economic factors originally influence on stock prices.

Therefore, after different methods that were examined, this was resulted, perhaps the Kalman filter _according to its computational features that were introduced in the previous section_ is capable for data fusion of both stock market analysis and forecasting more realistic price based on economic and time variables.

So in this section a structure with multi input-output was developed which had an extended Kalman filter in its heart, Structure designed to combine information of fundamental and technical analysis presented as below. According to the structure data are entered from two sides. The first part is some economic and financial parameters used by the fundamental analysis. These variables after moving to the structure put in the system model and then changes to price.

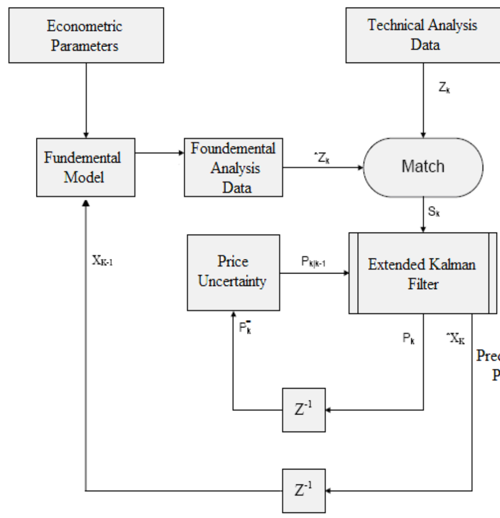

Figure 2. Data fusion structure based on Extended Kalman filter

Second part is the general trend of market prices _which is technical analysis data.

Kalman Filter can estimate and predict data with normal distribution. Therefore, it should be check if the stock price follows the normal distribution? According to the conducted research the answer is yes. Histogram on stock price of each company of Iran was drawn and then the normal curve was adjusted on the histogram and compared, the price changing of the shares follow from the normal distribution [14] [15].

There are two output results variable of the EKF part. One is x_k which is the predicted price of the share and used by a delay phase as the input economic model (if needed). The other is P which the covariance of the prediction error is. This is the only parameter that memorizes the effect of the past events for the next step of prediction.

Next section is about implementation of the model on the stock of an industrial company as case study.

IV. Model Implementation

To implement the model, 183 records of information for one year of Bahman industrial group were explored from

RAHAVARD NOVIN software. The following diagram shows trend of stock price.

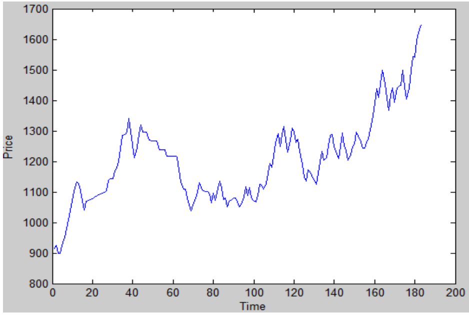

Figure 3. The stock price trend of an industrial company in Iran in one year

To implement the model, other parameters such as EPS, efficiency ratio, market efficiency and the coefficient β, were required for calculating the expected actual return and making economic model of stock price of Bahman group. After exploring data from the RAHAVARD NOVIN software, now, according to available information, the Gordon model can be assumed as a system model for the stock like Bahman group [17].

$$P_0 = \frac{E_0(1-b_i)}{k_i - br_i} \tag{20}$$

Where, P_0 is the price of the stock in the first year, b, the percent of held interest, *k*, is Expected rate of return on shareholders, r, investment efficiency rate, *br,* is dividend grow rate.

According to the parameters available in Fundamental Analysis, Gordon model was assumed as input state model of the system in this project.

$$x_k = f(x_{k-1}, u_{k-1}, w_{k-1}) = \frac{E_0(1-b_i)}{k_i - br_i} \tag{21}$$

Here, u are the financial and economic parameters of the stock market environment.

On the other hand the technical analysis of data related to stock price changes, as input prices obtained from previous occurred observations_ are calculated in the function h and then entered to the EKF.

$$\hat{z}_k = h(x_k, v_k) \tag{22}$$

As mentioned in the previous section the Kalman filter performance need the use of parameters like Q, R, W_k and V_k. These parameters are always set practically. Now let's calculate the statistical parameters by using available data. Q is covariance of process noise [16]

To calculate Q according to the parameters definition, an interval stock price changes should be calculated. For this purpose, in a specified period (one year) change in stock price Bahman Group is calculated by the formula.

$$\mathrm{e_i(n) = Price_i(n) - h_i(n)} \tag{23}$$

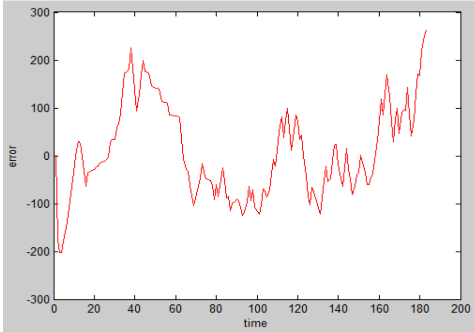

Figure 4. error between the real stocks prices and technical analysis predicted price

The error variance e_i is 0.777 which is Q, and according to $p(w) \sim N(0,Q)$, W_k is calculated 0.001.

R Is calculating, by_ considering the definition of the parameter_ the difference between real price and the state model price. According to below equation:

$$\mathrm{e_j(n) = f_j(n) - Price_j(n)} \tag{24}$$

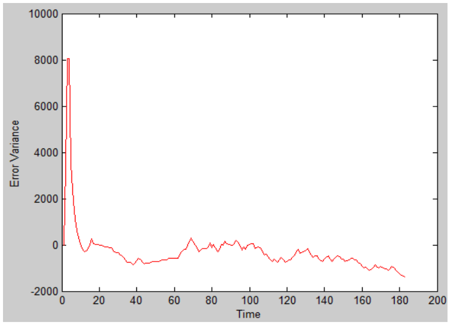

Figure 5. Changes of the error between actual price and model price

The error variance e_i is 0.1138 which is R, and according to $P(v) \sim N(0,R)$, v_k is calculated 0.01.

Now all the parameters are specified, Kalman filter can be developed to create the structure and values can be replaced [18]. First of all the parameter of the prediction phase should be calculated according to the state model.

$$A_k = f'(\hat{x}_k, u_k) = \begin{bmatrix} \frac{\partial f}{\partial E} & \frac{\partial f}{\partial b} & \frac{\partial f}{\partial k} & \frac{\partial f}{\partial br} \end{bmatrix} \tag{25}$$

$$A_k = \begin{bmatrix} \frac{\mathrm{1-b}}{\mathrm{k_n - br_n}} & -\frac{\mathrm{E_n}}{\mathrm{k_n - br_n}} & -\frac{E_n(1-b)}{(\mathrm{k_n - br_n})^2} & \frac{E_n(1-b)}{(\mathrm{k_n - br_n})^2} \end{bmatrix} \tag{26}$$

And then the observation model will be obtained by the simple regression calculation. This model (h) used as a method of technical analysis stock price prediction which is historical review of price trend.

$$y_k = h(x_k) + v_k \tag{27}$$

$$h_{\hat{x}_k}(n) = 0.0082\mathrm{n}^2 + 0.0467\mathrm{n} + 1101.1 \tag{28}$$

$$C_k = h'(\hat{x}_k) = 0.0164\mathrm{n} + 0.0467 \tag{29}$$

The extended Kalman filters formulas are according to equation (17), (18) and (19), mentioned in the previous section with the respect to the price changing's.

$$K_k = P_k^- C_k^T (C_k P_K^- C_k^T + R)^{-1} \quad (30)$$

$$\Delta\hat{x}_k = \Delta f(\hat{x}_k^-, u_k) + K_k(\Delta z_k - \Delta h(\hat{x}_k^-, 0)) \quad (31)$$

$$\hat{x}_k = \hat{x}_k^- + \Delta\hat{x}_k \quad (32)$$

$$P_k = A_k(I - K_k C_k H) P_k^- A^T{}_k + Q \quad (33)$$

V. RESULTS

After the implementation of the simulated structure code with MATLAB 7 following results were obtained. For example, after analyzing of one year data, by the structure, the price for the next day is predicted as follow.

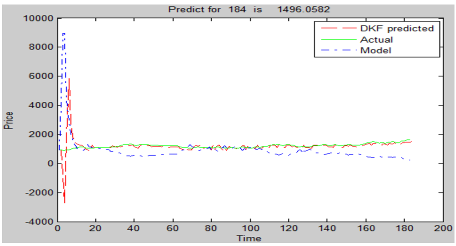

Figure 6. Prediction diagram of the next day stock price by EKF data fusion

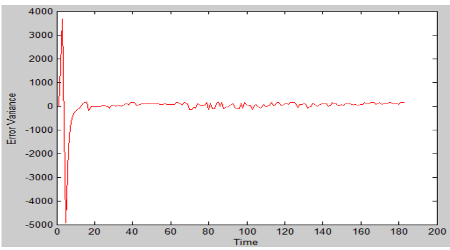

Figure 7. Changes of the error after passing through EKF structure

According to the above graph, reduction of the error in the predicted price in compare with the actual price, after leaving the Kalman filter, is observable. In continue, for more examination of the method, forecasting was done in the different horizon of time. Table I shows the result of the experiments.

TABLE I. STOCK PRICE PREDICTION IN DIFFERENT TIME HORIZON WITH EKF MODEL

Predict for	Occurred price	Predicted price	MSE	MAE
Next day	1587	1496	25.9	156.76
Next week	1587	1681	62.7	53.9
Next month	1587	1584	25.7	39.6

Also, according to the mission of the Kalman filter, which is "Minimizing the estimated error covariance" in equation (19),

$$\lim_{P^-{}_k \to 0} K_k = 0 \quad (34)$$

Figure 8 shows this fact in the implemented method.

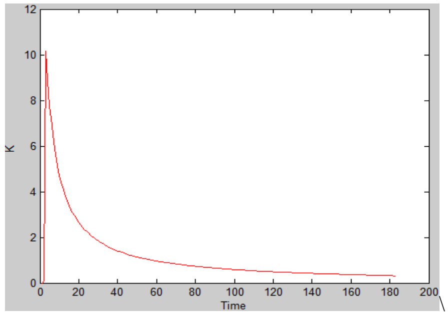

Figure 8. Minimization of the error covariance

VI. CONCLUSIONS

This article was trying to predict stock prices of the stock market in the short term, relying on a new fusion method called extended Kalman filter that enables to use fundamental analysis of stock price with technical analysis.

In the presented structure, simultaneous analysis of technical data (with the indicators) with fundamental analysis data (with financial variables that include general stock price model) were entered to prediction-correction filter named extended Kalman filter. Then, after statistical and possibilities calculation, a new stock price will be presented as the algorithm predicted price to the user.

As the graphs in Figure 6 and 7 shows, it is clear that the error of prediction reduced much after passing through the filter, which shows the proper efficiency of the presented model in this research.

Data in the table II, compare the calculated error values predicted in the three offered methods for the same time horizon of the same company. For better comparison, scientific indicators such as mean square error (MSE) and mean absolute error (MAE) was measured for each method to show the success of the algorithm in this paper.

TABLE II. ERROR COMPARISON IN DIFFERENT METHOD OF PREDICTION

Methods	MSE	MAE
Regression method	92.44	79.44
Artificial neural networks	40.09	154.46
Extended Kalman filter data fusion	25.90	156.76

REFERENCES

[1] Pring, Martin J. "Introduction to Technical Analysis;", Mc Grow Hill.1998.

[2] Plummer. T; Forecasting Financial markets: A Technical Analysis and the Dynamic of price; John Wiley; N.Y. 1991.

[3] Fugence,F. Fama.," Efficient Capital Market: A Review of Theory and Enprical Work", The Journal of Finance, No.2, May1970, PP.383-417.

[4] Donald B., Keim.," Size-Related Anomalies and Stock Return Seasonality: Further Empirical Evidence", Journal of Financial Economics 12 , 1983.

[5] Namazi. M; Z. Shoushtarian; “A Review of Test on the weak form of Efficient market hypothesis”; Journal of Financial Research; Vol. 13; No 11 and 12,1996.

[6] Todd F., M.. Correa, A..,”Gaussian Process Regression Models for Predicting Stock Trends”, MIT technical report , 2007.

[7] Abbasi, E., Abouec, A.,” Stock Price Forecast by Using Neuro-Fuzzy

[8] Inference System”, International Journal of Business and Economic Sciences 1:3 2009

[9] Qi, M. & G.S. Maddala, “Economic Factors and The Stock Market: A New Perspective;”, Journal of Forecasting18, pp: 151-166. 1999.

[10] Kalman,R.E,“A New Approach to Linear Filtering and Prediction Problems”, Transaction of the ASME-Journal of Basic Engineering (series D):pp: 35-45. 1960 .

[11] McMillan, D. G. ‘Non-Linear Predictability of Stock Market Returns: Evidence from Non-Parametric and Threshold Models’, International Review of Economics and Finance, Vol.10, pp. 353-368. 2001.

[12] Simon, D. “Optimal State Estimation: Kalman, H∞ , and Nonlinear Approaches”, New York:John Wiley & Sons, 2006.

[13] Welch, G., Bishop,G. ”An Introduction to the Kalman filetr”. by ACM, Inc. 2001.

[14] Haykin, Simon, ”Kalman Filtering And Neural Networks”, AWiley&Science Publication,2001, ISBN 0-471-36998-5.

[15] Namazi, Mohammad, Shvshtryan, Zkyeh., "Stock Market Performance Analysis of Iran. Stock Exchange Market", Journal of Financial Research and Scientific Research, Second Year, No. 7 and 8, pages 82-104. 1994.

[16] Francis, J.C., “Investment: Analysis and management”,New York, 1972, McGraw-Hill.

[17] Huang , Shian-Chang, “Online option price forecasting by using unscented Kalman filters and support vector machines”, Journal of Expert Systems with Applications 34 , 2008, 2819–2825.

[18] Gordon, Myron J. . "Dividends, Earnings and Stock Prices". Review of Economics and Statistics (The MIT Press) 41 (2): 99–105, 1959.

[19] S.Maybeck,P.”Stochastic models, estimation and control”, Harcourt Brace Jovanovich Publisher , Academiv Press. 1979..

Hassan Haleh, born on February of 1968 at Tehran -Iran. He got his Phd. in Production Engineering from Gunma University-Japan in 2004.

He has been assistant professor of Qazvin Azad University since 2005. His fields of research are Multi Criteria Decision Making, Automation, Project management and robotics. Also he has been the advisor of some financial and stock market analysis thesis. He is teaching MCDM, OR and automation.

Dr. Haleh has been dean of Industrial and mechanical engineering faculty at Qazvin Azad University since 2008.

Beitollah Akbari Moghaddam, born on August of 1969 at Rasht- Iran. He got his Phd. in mathematical economics from the university of Mazandaran.

He has been assistant professor at Qazvin Azad University since 1996. He has been working on computable general equilibrium, calculus of variation applied in economics. He is teaching Micro, Macro and management economics.

Dr. Akbari Moghaddam has been the financial vice president of QIAU since 2010.

Saeed Ebrahimijam, born on October of 1983 at Zanjan-Iran. He got his BSc. in computer engineering in 2007 and then graduated from MSc. in Industrial engineering, system management and productivity from Qazvin Azad University.

He is a lecturer at Qazvin Azad University and has been working on intelligent algorithm, machine learning and computational method in robotics, and now he is expanding the usage of such methods in economical and financial subjects. He is IT project management, principals of management ssand entrepreneurship.

Mr. Ebrahimijam has been researcher at mechatronics research laboratory of QIAU since 2002 and leader of MRL middle size soccer robot team. He is member of technical committee and organization committee of robocup federation since 2009