JME

Journal of Mining & Environment,
Vol.9, No.2, 2018, 349-360.
DOI: 10.22044/jme.2017.6195.1435

Forecasting copper price using gene expression programming

H. Dehghani

Department of Mining Engineering, Hamedan University of Technology, Hamedan, Iran

Received 8 September 2017; received in revised form 28 November 2017; accepted 29 November 2017
Corresponding author: dehghani@hut.ac.ir (H. Dehghani).

Abstract

Forecasting the prices of metals is important in many aspects of economics. Metal prices are also vital variables in financial models for revenue evaluation, which forms the basis of an effective payment regime using resource policymakers. According to the severe changes of the metal prices in the recent years, the classic estimation methods cannot correctly estimate the volatility. In order to solve this problem, it is necessary to use the artificial algorithms, which have a good ability to predict the volatility of various phenomena. In the present work, the gene expression programming (GEP) method was used to predict the copper price volatility. In order to understand the ability of this method, the results obtained were compared with the other classical prediction methods. The results indicated that the GEP method was much better than the time series and multivariate regression methods in terms of the prediction accuracy.

Keywords: *Copper Price, Gene Expression Programming, Forecasting, Time Series.*

1. Introduction

Copper is one of the most important industrial metals, which plays vital roles in various aspects in today's economies. Similar to crude oil, gold, and other commodity markets, the future market is the main pricing and trading market for copper [1]. Copper future price has a marked impact on some metal prices [2]. On the other hand, for some countries such as Chile and Zambia, whose economy relies extensively on copper production, fluctuations in copper price is very important [3].

Therefore, knowing the copper price changes may play an important role in making the correct decisions for applying the administrative options for extending or restricting the mining activities via mining project managers and shareholders. Significant volatilities in the copper price, especially in the recent years, has led to the classic prediction approaches that do not have the ability to estimate the price changes. Hence, numerous researchers have tried to predict the mineral prices using artificial methods. Xie et al. have proposed a new method for crude oil price prediction based on a support vector machine (SVM) model. They compared their model with the other ones, which were developed using artificial neural networks (ANNs) and genetic algorithm (GA). The results obtained show that like ANN and GA, SVM is a capable method for forecasting the crude oil price [4]. Hadavandi et al. have developed a time series model for gold price and exchange rate forecasting based on particle swarm optimization (PSO) [5]. Dehghani and Ataee-pour have predicted the copper price using binomial tree [6]. Dehghani et al. have estimated the price and operating cost in a copper mine using multi-dimensional binomial tree [7]. Kriechbaumer et al. have used an improved combined wavelet-autoregressive integrated moving average (ARIMA) to forecast the monthly price of aluminum, copper, lead, and zinc [8]. Li and Li have studied the volatility of the copper price using time series functions [1]. Lasheras et al. have used the ARIMA and ANNs methods in order to predict the copper spot price in New York Commodity Exchange (COMEX). The results of this research work show that the estimation error of the neural network is always less than the time series model [3]. Chen et al. have investigated the

changes in various metals prices using the grey wave forecasting method [9]. Mostafaa and El-Masry have predicted the crude oil price using gene expression programming (GEP) and ANNs [10]. Liu and li have forecasted the gold price and analyzed the related influential factors based on random forest [11]. Liu et al. have predicted the copper price using decision tree learning. Their model forecast the copper price using price volatility of several materials such as crude oil, gold, and silver [12]. Dehghani and Bogdanovic have proposed a new model based on the time series functions and bat algorithm [13].
According to the above-mentioned points and importance of copper price volatility, the current research work tries to estimate the copper price using GEP.

2. Gene expression programming (GEP)

GEP is a new, popular, and evolutionary technique (EC) that deals with complex types of problems through the use of a linear tree representation [14]. In fact, GEP was introduced as a reaction to the complexity that GP experiences with tree structures, and the difficulty that other linear representations of programs experience in ensuring the validity of their evolved structures. GEP is able to create trees indirectly, by encoding the mass vectors of symbols and translating them into trees only in order to evaluate their fitness. This allows simple genetic operators, as found in GAs, while evolving complex and expressive trees, as GP does. It is also justified biologically in what Ferreira calls the “phenotype barrier”, where the genotype must be expressed as a more complex structure in order to have an effect on the environment [15]. The EC techniques are useful when the search space is large and complex, and solutions are ill-defined apriori [16]. The EC techniques are based upon the Darwinian evolution principle, which suggests that populations evolve through inheritance where a concept of fitness reflects the population's ability to survive. GEP was first introduced to the GP community by Ferreira [17, 18]. Thus it is the most recent development in the field of artificial evolutionary systems [18]. GEP starts with allocating a fixed length chromosome to the randomly generated initial population. Then the chromosomes are explicitly expressed, and each individual's fitness is evaluated. The individuals with high fitness are selected to improve the solution. This process is iterated for a pre-specified number of generations or until an “optimal” solution has been found. Figure 1 provides a flowchart of the GEP structure.

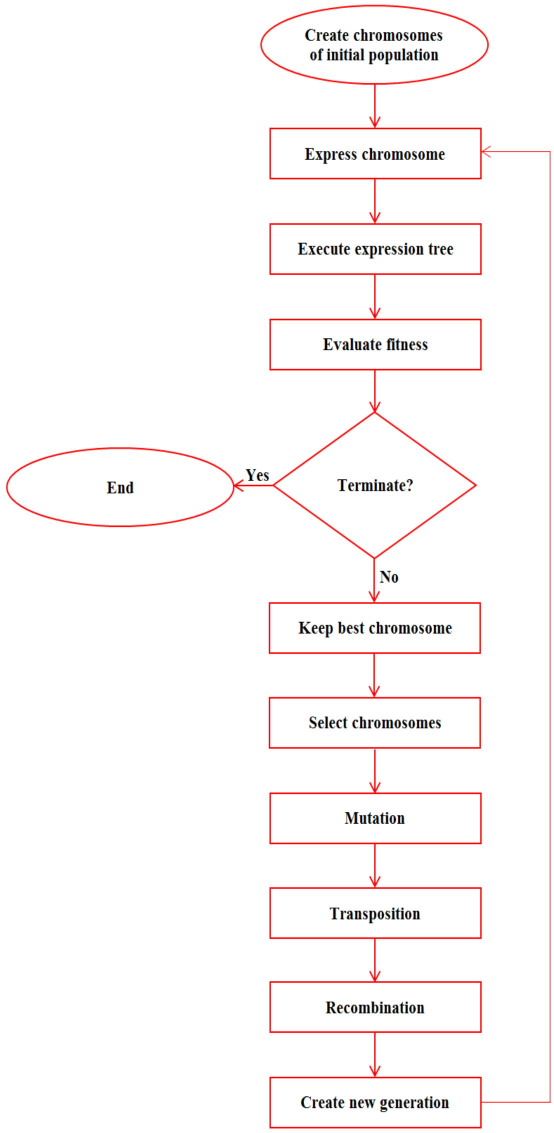


Figure 1. Flowchart of GEP.

3. Methodology

Many factors may contribute to the fluctuation of copper prices. Charlot and Marimoutou have mentioned that the price of copper may relate to other metal prices in the market, as alternatives [19]. For example, when other metal prices fall down, the copper price is likely to follow the same trend. Chang, et al. believed that copper price could be related to some variables such as lean hogs price and coffee price in an unanticipated way [20]. Joseph and Kundig have shown that the copper price is likely to depend upon energy cost such as the prices of crude oil and natural gas due to the connection to the production of copper [21]. As Gargano and Timmermann mentioned in their research work, the copper price could be related to the demand and consumption in industries, which turns to be affected by the general economy environment [22]. Therefore, the copper price might be a function of generic economic indicators such as the NIKKEI and Dow Jones indices. Since copper producing countries trade their products on world markets with dollar denominations, the exchange rate between supplier and demander currency might be correlated to copper price.

Based on the above reasoning, and subjected to the data availability, for investing the effect of the above-mentioned parameters on the copper price volatilities, at the first step, the affecting parameters were divided into five major groups, i.e. energy price, metal price, exchange rate, global stock market changes, and crude price. At the second step, the datasets were gathered from 2009 to 2016. All the data were downloaded from http://www.investing.com. Then the copper price changes were forecasted using three methods, i.e GEP, multivariate regression, and time series function. Finally, the results obtained were compared together. Figure 2 shows the methodology of the current research work.

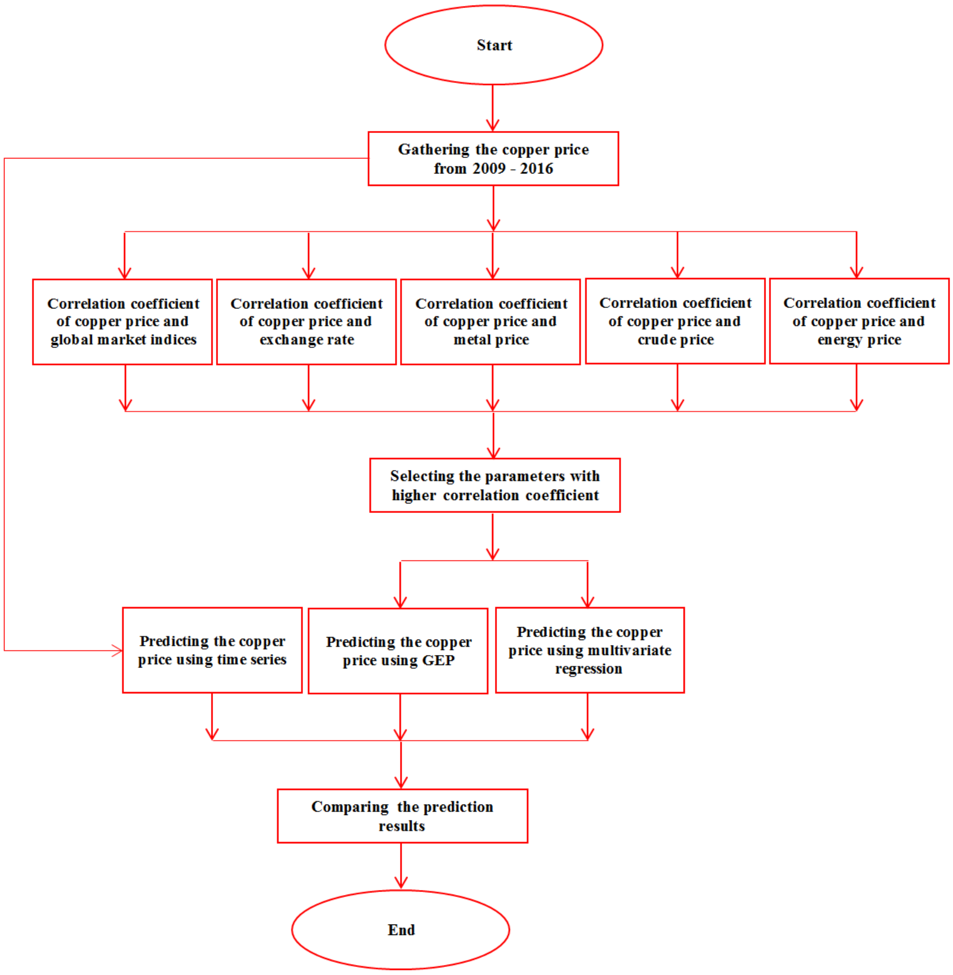


Figure 2. Flowchart of copper price prediction.

3.1. Correlation between copper price and other commodity prices

Pearson cross-correlation coefficients were used to quantify the observed qualitative correlations between copper and the other selected variables. The Pearson cross-correlation coefficient measures the linear correlation, and therefore, the degree of linear dependence, between two variables, giving a value between +1 and −1 [23]. A coefficient of +1 is total positive correlation, 0 is no correlation, and −1 is total negative correlation [23]. The Pearson cross-correlation coefficient is defined as follows:

$$r_{i,c} = \frac{\operatorname{cov}(P_i, P_c)}{\sigma_{P_c} \times \sigma_{P_i}} \tag{1}$$

where $r_{i,c}$ is the Pearson cross-correlation coefficient between parameter i and copper price,

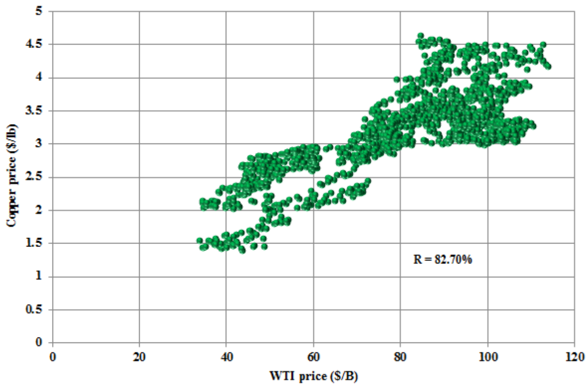
σ_{P_c} and σ_{P_i} are the standard deviations of the copper price and value of parameter i, respectively, and the covariance between the copper price and value of parameter i is calculated as below:

$$\operatorname{cov}(P_i, P_c) = \frac{1}{n}\sum_{j=1}^{n}\left(P_{cj} - \bar{P}_c\right)\left(P_{ij} - \bar{P}_i\right) \qquad (2)$$

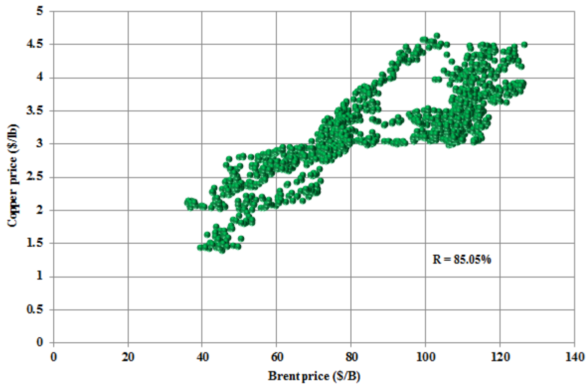
where P_{cj} is the price of copper in day j, P_{ij} is the value of variable in day j, $\bar{P}_c$ is the average of the copper prices, $\bar{P}_{ij}$ is the average of the value of the variable, and n is the number of datasets.

3.1.1. Copper price versus energy price

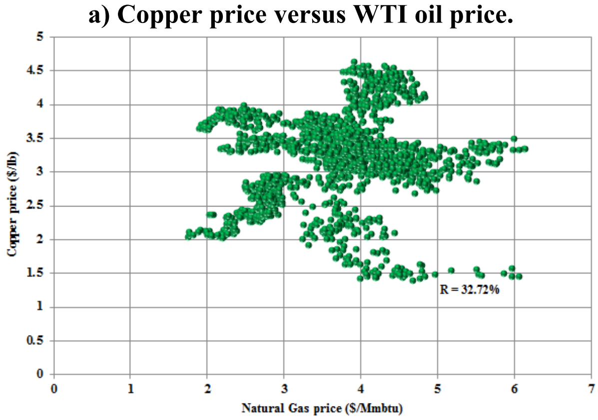
In order to find the effect of the energy price changes on the copper price volatilities, the price of the main energies such as Brent crude oil, OPEC crude oil, West Texas Intermediate (WTI), coal, and natural gas were gathered. The coefficient of correlation of the copper price and energy prices from 2009 to 2015 is shown in Figure 3. It is obvious that Brent crude oil price and natural gas price had the highest and lowest correlation with copper price, respectively. The small correlation coefficient for the natural gas might indicate that it plays a less important role than crude oil in the production process of copper [21].



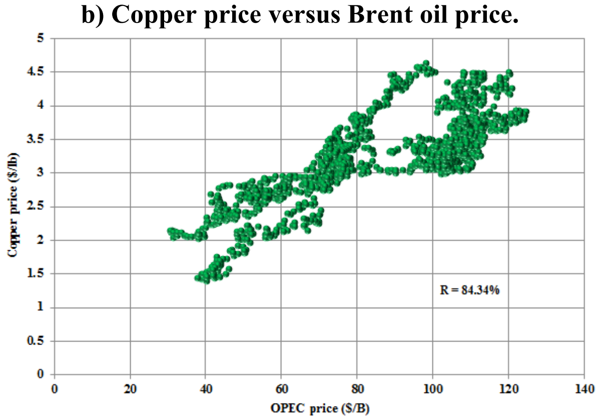
a) Copper price versus WTI oil price.



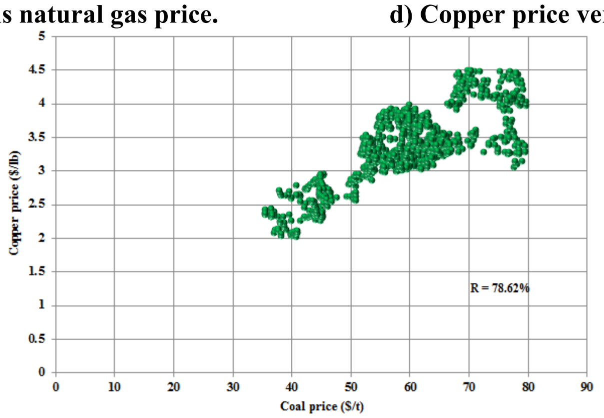
b) Copper price versus Brent oil price.



c) Copper price versus natural gas price.



d) Copper price versus OPEC oil price.

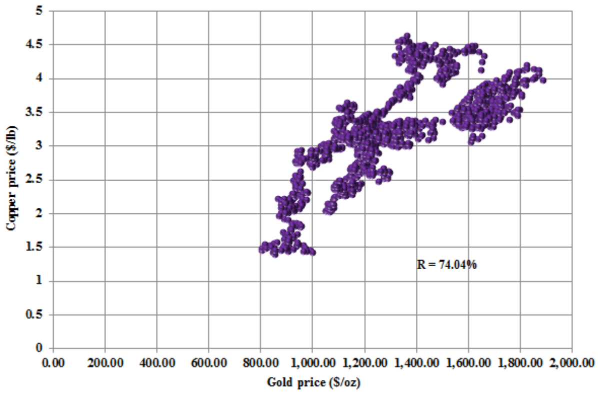


e) Copper price versus coal price.

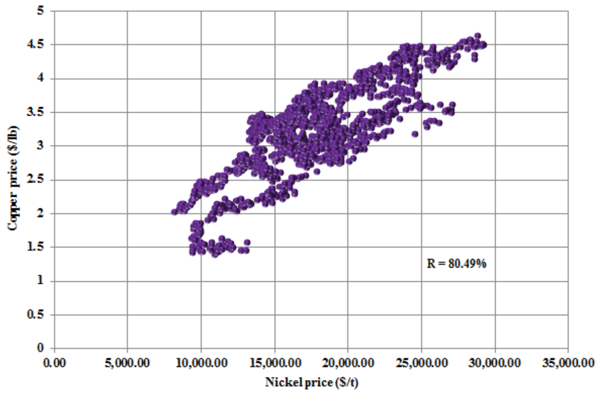
Figure 3. Correlation between copper price and energy prices.

3.1.2. Copper price versus metal prices

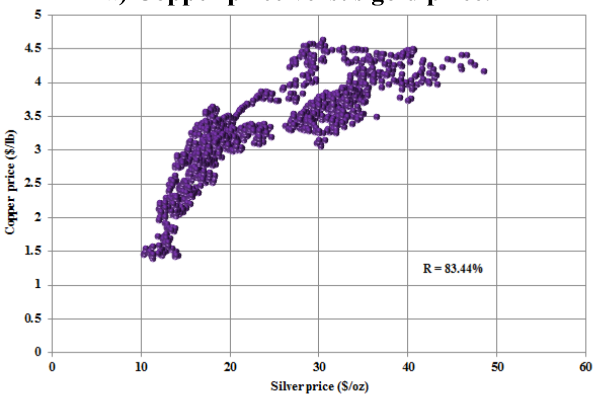
In order to study the effect of the metal price changes on the copper price volatilities, the price of some metals such as gold, silver, zinc, nickel, and aluminum were gathered. The coefficient of correlation of the copper price and metal prices from 2009 to 2015 is shown in Figure 4. According to this figure, aluminum price and zinc price had the highest and lowest correlation with copper price, respectively.



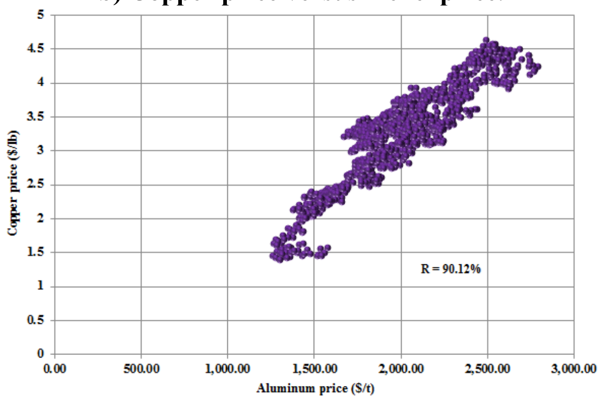
a) Copper price versus gold price.



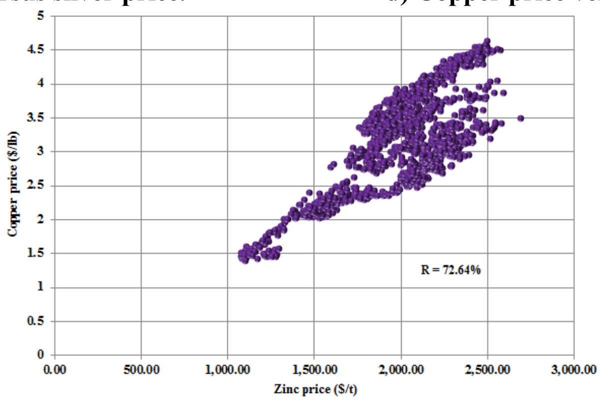
b) Copper price versus nickel price.



c) Copper price versus silver price.



d) Copper price versus aluminum price.

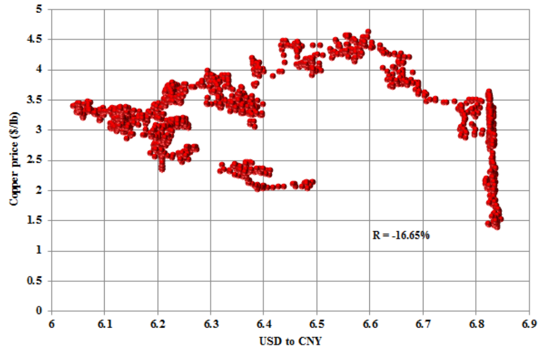


e) Copper price versus zinc price.

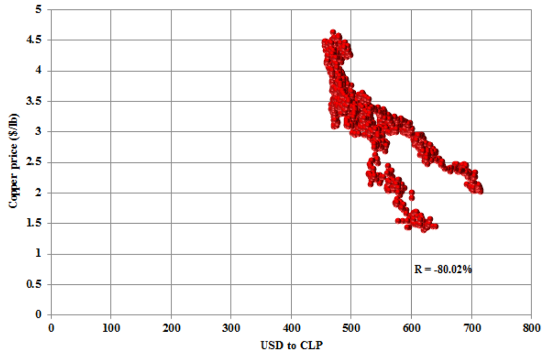
Figure 4. Correlation between copper price and metal prices.

3.1.3. Copper price versus exchange rates

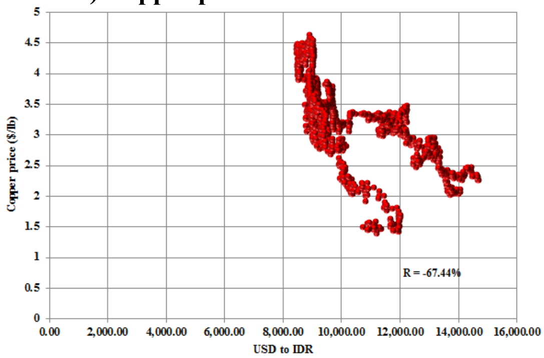
In order to investigate the effect of the exchange rate changes on the copper price volatilities, the exchange rates of some copper supplier and demander countries were gathered. The currency of Chile and China, as the biggest producers of the copper, and China, European countries, and USA, as the biggest consumers of the copper, i.e. USD (United States Dollar), CLP (Chilean Peso), EURO (European Currency Unit), CNY (Chinese Yuan) and IDR (Indonesian Rupiah) were selected. The coefficient of correlation of the copper price and exchange rates from 2009 to 2015 is shown in Figure 5. According to this figure, USD/CLP and USD/CNY had the highest and lowest correlation with copper price, respectively.



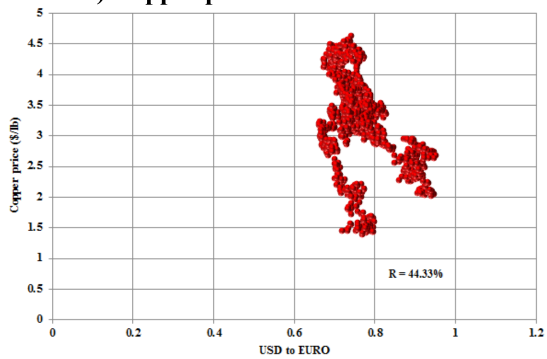
a) Copper price versus USD/CNY.



b) Copper price versus USD/CLP.



c) Copper price versus USD/IDR.

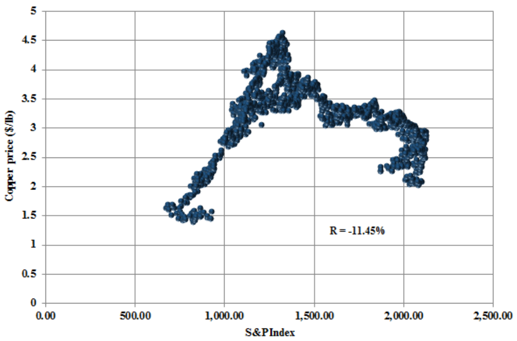


d) Copper price versus USD/EURO.

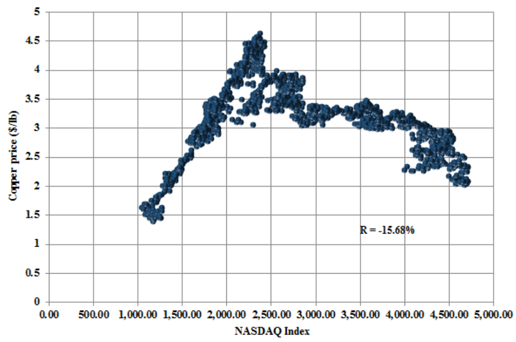
Figure 5. Correlation between copper price and exchange rate.

3.1.4. Copper price versus global stock market changes

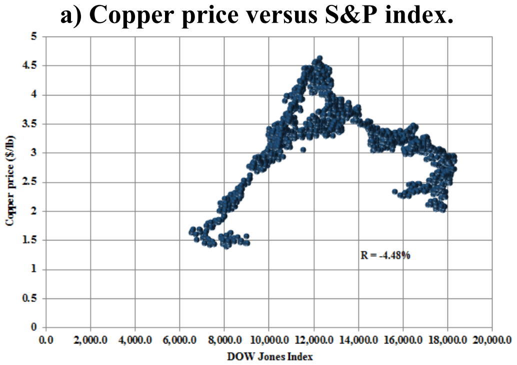
In order to find the effect of the global stock market changes on the copper price volatilities, the indices of some global stock markets such as S&P, NASDAQ, Dow Jones, and NIKKEI were gathered. The coefficient of correlation of the copper price and global stock market changes from 2009 to 2015 is shown in Figure 6. According to this figure, there is no significant correlation between these indices and copper price.



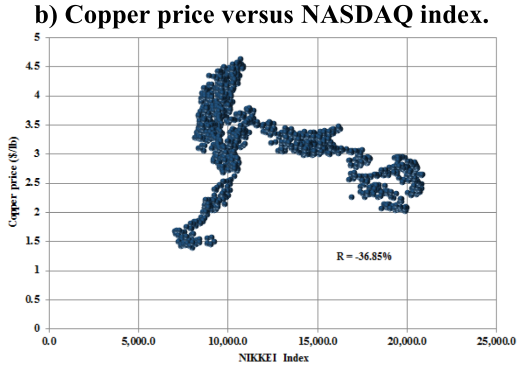
a) Copper price versus S&P index.



b) Copper price versus NASDAQ index.



c) Copper price versus DOW Jones index.

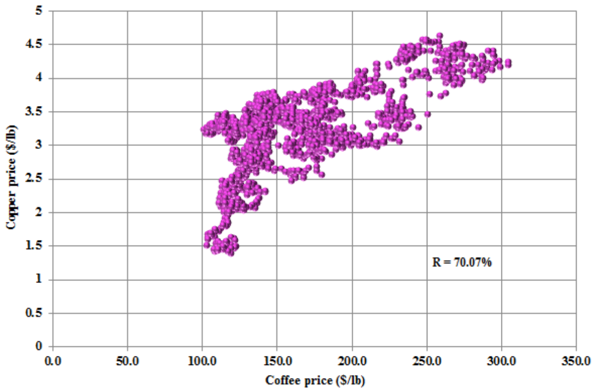


d) Copper price versus NIKKEI index.

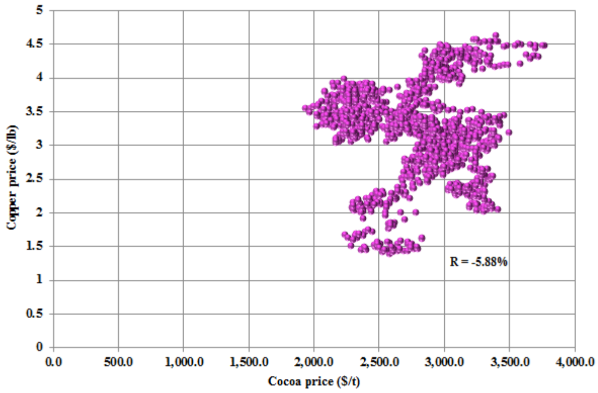
Figure 6. Correlation between copper price and indices.

3.1.5. Copper price versus crop price

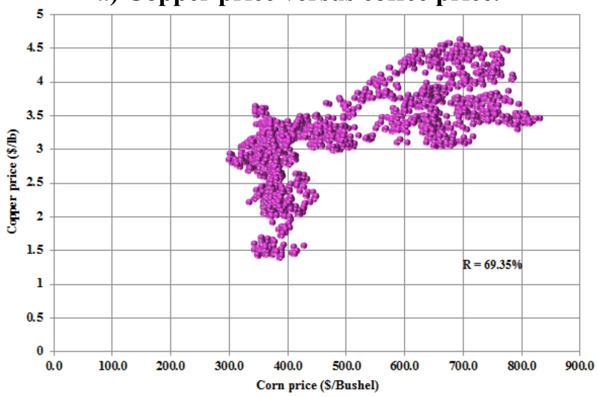
In order to investigate the effect of the crop price changes on the copper price volatilities, the price of some crops such as wheat, corn, coffee, and cocoa were gathered. The coefficient of correlation of the copper price and crop prices from 2009 to 2015 is shown in Figure 7. It is obvious that the coffee price and cocoa price had the highest and lowest correlation with the copper price, respectively.



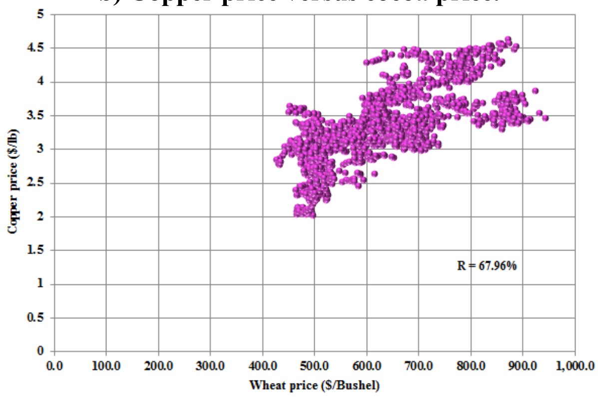
a) Copper price versus coffee price.



b) Copper price versus cocoa price.



c) Copper price versus corn price.



d) Copper price versus wheat price.

Figure 7. Correlation between copper price and food price.

4. Copper price prediction

Using Figures 3-7, the most correlated parameters with the copper price were selected. Figure 8 shows the coefficient of correlation between the considered parameters and the copper price. It is assumed that the copper price is a function of the important variables, the score of which was at least 80%, i.e. silver price, nickel price, aluminum price, OPEC crude oil price, WTI crude oil price, BRENT crude oil price, and USD to CLP exchange rate.

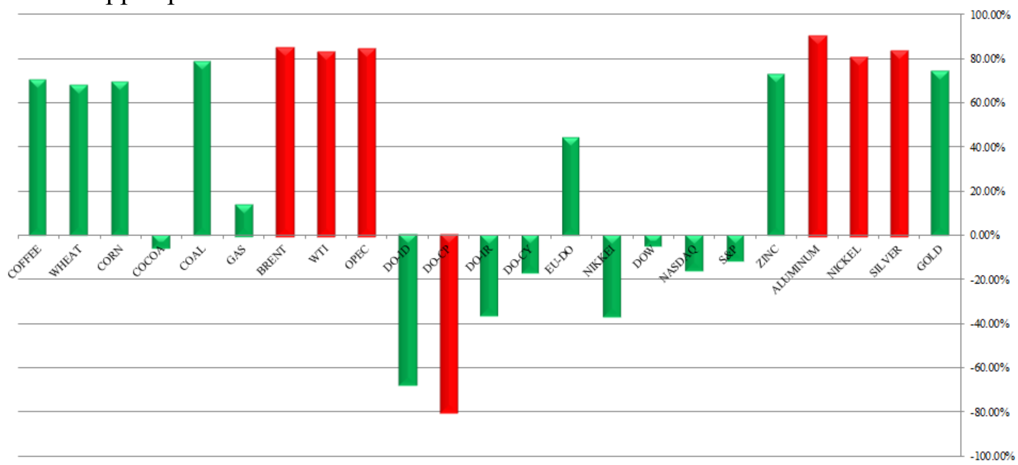


Figure 8. Pearson's correlation coefficients between copper price and selected affecting factors.

4.1. Gene expression programming (GEP) in this study

Since the copper price time series appears to be non-linear and non-stationary, the GEP model can be used. This technique has been selected because it is able to perform non-linear modeling and adaptation. It also does not assume a priori any functional form of the time series analyzed [24]. In order to conduct the analysis, the silver price, nickel price, aluminum price, OPEC crude oil price, WTI crude oil price, BRENT crude oil price, and USD to CLP exchange rate were assumed as the input and the copper price was assumed as the output. For developing the model, the datasets were partitioned into a training set (from 2009 to 2015) and a test set (from 2016 to 2017). GeneXpro 5.0 software package was used to conduct the analysis because this software has extensive GEP and soft computing tools.

The GEP modeling includes five major stages [25]. The first one is to select the fitness function. In this work, an R-squared fitness function was used, which is very useful in multivariate regression applications since it selects the model with the highest explanatory power. The second stage involves the selection of a set of terminals (T) and a set of functions (F) in order to form chromosomes. In this work, terminals were considered $T = \{P_i(t-1), P_i(t-2), ...\}$, where $P_i(t-1)$ is the value of the variable i in time $(t-1)$. In this work, the four major arithmetic operators along with several other functions were used: $F = \{+, -, \times, /, power, \ln(x), square, ...\}$.

The third stage involves the selection of the chromosomal architecture. In this work, we selected a chromosome with a head length = 8 and the number of genes per chromosome was set to 7. The fourth stage encompasses the selection of the type of linking function. The addition function was used to link the sub-expression trees. Finally, in the last stage, a set of genetic operators and rates have to be selected. In this work, in order to increase fitness, a combination of all possible genetic operators such as mutation, transposition, and cross-over were selected. The genetic operators and rates were chosen using the trial and error method. Table 1 shows the GEP parameters used in this work.

The best fitness obtained using the GEP specifications mentioned above was 0.931 for predicting data (2009-2015) and 0.801 for validating data (2016-2017), and the corresponding expression trees are shown in Figure 9.

Based on the statistical results presented above, the GEP technique appears to be very accurate in predicting copper prices. Figure 10 shows the real series versus the predicted series from 2016 to 2017.

Table 1. GEP parameters.

Parameter	Value	Parameter	Value
Generations required to train model	1000	Gene size	58
Complexity of model before simplification	60	Head size	8
Complexity of model after simplification	26	Tail size	20
Generations required for simplification	25	Linking function	Addition
Inversion rate	0.1	Mutation rate	0.045
IS transportation rate	0.1	Chromosome length	30
RIS transportation rate	0.1	Number of genes	7
One-point recombination rate	0.3	Gene recombination rate	0.1
Two-point recombination rate	0.3	Gene transportation rate	0.1

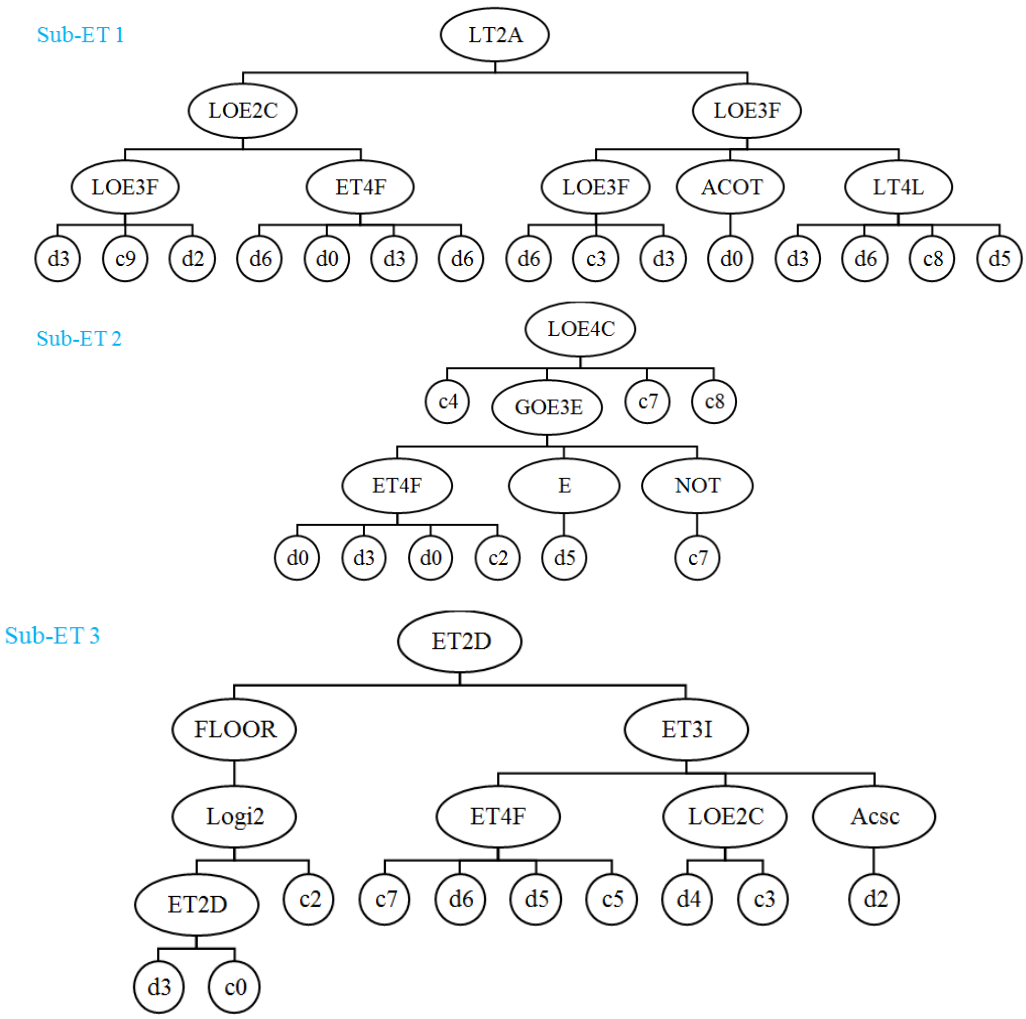


Figure 9. Model expression tree.

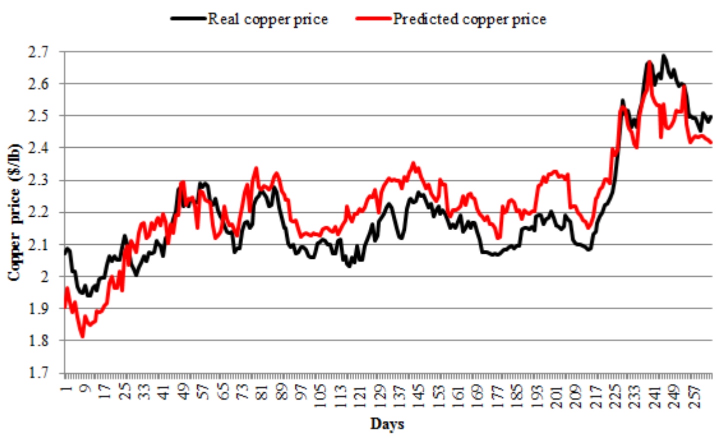


Figure 10. Real copper price versus predicted copper price.

4.2. Multivariate regression

In order to compare the results of the developed model with other forecasting models, a multivariate regression model was developed using the input and output data. For preparing this model, SPSS 20.0 was used. Equation 3 shows the multivariate regression model.

$$P_{copper} = 0.331756 + 0.008012 \times P_{silver} + 0.000994 \times P_{WTI} + 0.028878 \times P_{brent} - 0.00094 \times P_{USD/CLP} - 0.021347 \times P_{OPEC} + 0.00000163 \times P_{nickel} + 0.001191 \times P_{alu\ minum} \quad (3)$$

Figure 11 shows the correlation coefficient of the presented model.

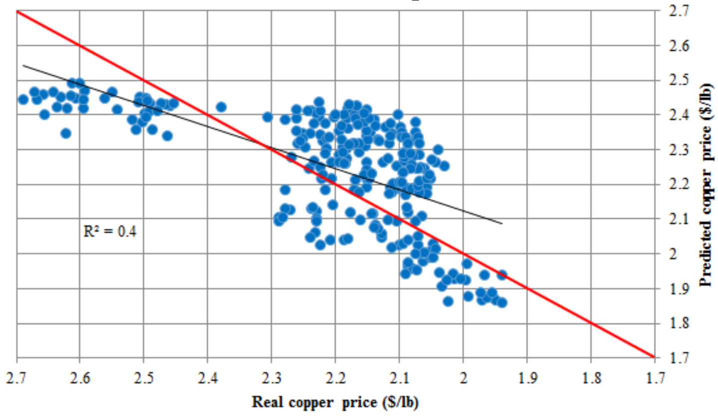


Figure 11. Comparison between real and predicted copper price.

4.3. Time series function

In order to develop the copper price volatility time series estimation model, the copper price historical datasets were gathered from 2009 to 2016. The datasets from 2009 to 2015 were used for training the models, and the 2016-2017 datasets were used for validating the models. @risk (Ver. 7.5) was applied for estimating the copper price changes using the time series. In order to achieve this purpose, all functions were used. It was observed that BMMR and GARCH were the best and worst copper price estimation functions, respectively. The estimation parameters (μ,σ,γ) of BMMR function were 3.285, 0.109, and 0.019, respectively. Figure 12 shows the distribution of the simulated correlation coefficient of the real and predicted copper price. The number of iterations was considered to be 5000. The mean correlation coefficient for the proposed model was 0.37. Although the minimum and maximum correlation coefficient were -0.80 and 0.87, at the 90 percent confidence level, the correlation coefficient was between -0.25 and 0.72.

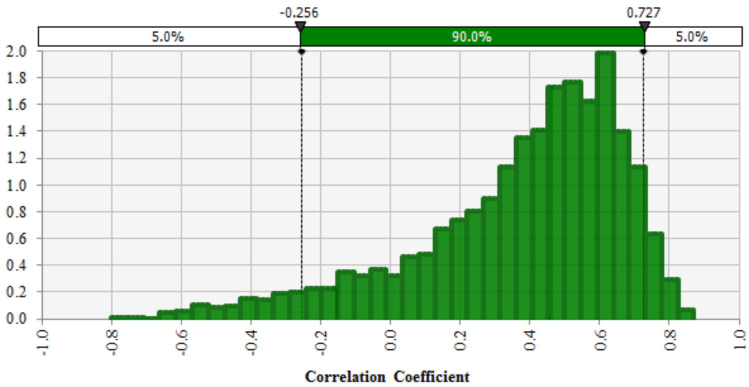


Figure 12. Simulation of correlation coefficient.

5. Discussion

Comparison between the forecasting models is presented in Table 2. According to this table, the results obtained from GEP are better than those from the multivariate regression and time series methods. It is because of the great ability of GEP for using the several mathematical functions in the forecasting process. Also using the multi-parameters has increased the accuracy of estimation.

Table 2. Comparison between predicting models.

	Correlation coefficient	R-Squared	RMSE	MSE
Gene expression programming	0.80	0.64	0.17	0.03
Multivariate regression	0.63	0.40	0.18	0.03
Time series functions	0.37	0.14	0.55	0.37

6. Conclusions

In this work, the copper price was estimated using multivariate regression, time series, and gene expression programming (GEP). The aim of this research work was to find the ability of GEP in order to estimate the commodity price. According to this, in the first stage, the best estimation models of time series functions and multivariate regression were selected for copper price volatility prediction. After that, the results obtained were compared with the GEP estimation results. The most important time series functions were used for estimating the copper price changes. Among them, the BMMR time series function with a mean RMSE of 0.55 presented the best estimation. Also RMSE of the multivariate regression model was 0.18.

The results obtained for GEP provide remarkable improvements. The results have a much higher accuracy and, unlike BMMR and multivariate regression models, they are closer to the reality. Based on the foregoing facts, it can be confirmed that the suggested methodology can be successfully applied for prediction of the copper price volatility.

So far this methodology has only been applied for the copper price volatility prediction but it is obvious that the proposed model can be used for solving a large scale of prediction problems in various fields such as mining, management, economy, and production.

References

[1]. Li, G. and Li, Y. (2015). Forecasting copper futures volatility under model uncertainty. Resources Policy. 46:167-176.

[2]. Morales, L. and Andreosso-O'Callaghan, B. (2011). Comparative analysis on the effects of the Asian and global financial crises on precious metal markets. Res. Int. Bus. Financ. 25 (2): 203-227.

[3]. Lasheras, F.S., Cos Juez, F.J.D., Sanchez, A.S., Krzemien, A. and Fernandez, P.R. (2015). Forecasting the COMEX copper spot price by means of neural networks and ARIMA models. Resources Policy. 45: 37-43

[4]. Xie, W., Yu, L., Xu, S. and Wang, S. (2006). A new method for crude oil price forecasting based on support vector machines. International Conference Computational Science (part IV). pp. 444-451.

[5]. Hadavandi, E., Ghanbari, A. and Abbasian, S. (2010). Developing a Time Series Model Based On Particle Swarm Optimization for Gold Price Forecasting. Third International Conference on Business Intelligence and Financial Engineering. pp. 337-340.

[6]. Dehghani, H. and Ataee-pour, M. (2012). Determination of the effect of operating cost uncertainty on mining project evaluation. Recourse Policy. 37: 109-117.

[7]. Dehghani, H., Ataee-pour, M. and Esfahanipour, A. (2014). Evaluation of the mining projects under economic uncertainties using multidimensional binomial tree. Resources Policy. 39: 124-133.

[8]. Kriechbaumer, T., Angus, A., Parsons, D. and Casado, M.R. (2014). An improved wavelet- ARIMA approach for forecasting metal prices. Resource Policy. 39: 32-41.

[9]. Chen, Y., Zou, Y., Zhou, Y. and Zhang, C. (2016). Multi-step-ahead Crude Oil Price Forecasting based on Grey Wave Forecasting Method. Procedia Computer Science. 91: 1050-1056.

[10]. Mostafaa, M. and El-Masry, A. (2016). Oil price forecasting using gene expression programming and artificial neural networks. Economic Modelling. 54: 40-53.

[11]. Liu, D. and Li, Z. (2017). Gold Price Forecasting and Related Influence Factors Analysis Based on Random Forest. Proceedings of the Tenth International Conference on Management Science and Engineering Management. Advances in Intelligent Systems and Computing. Vol 502. Springer.

[12]. Liu, C., Hu, Z., Li, Y. and Liu, S. (2017). Forecasting copper prices by decision tree learning. Resources Policy. 52: 427-434.

[13]. Dehghani, H. and Bogdanovic, D. (2017). Copper price estimation using bat algorithm. Resources Policy. https://doi.org/10.1016/j.resourpol.2017.10.015.

[14]. Ryan, N. and Hibler, D. (2011). Robust gene expression programming. Procedia computer science. 6: 165-170.

[15]. Ferreira, C. (2006). Gene Expression Programming: Mathematical Modelling by an Artificial Intelligence. Vol. 21. Springer. San Francisco.

[16]. Chen, M. and Huang, S. (2003). Credit scoring and rejected instances reassigning through evolutionary computation techniques. Expert Syst. Appl. 24: 433-441.

[17]. Ferreira, C. (2001). Gene expression programming: a new adaptive algorithm for solving problems. Complex Sys. 13: 87-129.

[18]. Ferreira, C. (2004). Gene expression programming and the evolution of computer programs. In: de Castro, L.N., Von Zuben, F.J. (Eds.), Recent Developments in Biologically Inspired Computing. Idea Group Publishing. pp. 82-103.

[19]. Charlot, P. and Marimoutou, V. (2014). On the relationship between the prices of oil and the precious metals revisiting with a multivariate regime-switching decision tree. Energy Econ. 44: 456-467.

[20]. Chang, L., Zhenhua, H., Yan, L. and Shaojun, L. (2017). Forecasting copper prices by decision tree learning. Resources Policy. 52: 427-434.

[21]. Joseph, G. and Kundig, K.J.A. (1999). Copper: Its Trade, Manufacture, Use, and Environmental Status.

[22]. Gargano, A. and Timmermann, A. (2014). Forecasting commodity price indexes using macroeconomic and financial predictors. Int. J. Forecast. 30 (3): 825-843.

[23]. Pearson, K. (1895). Notes on regression and inheritance in the case of two parents. In: Proceedings of the Royal Society of London 58. pp. 240-242.

[24]. Teodorescu, I. (2006). Gene expression programming approach to event selection in high energy physics. IEEE Trans. Nucl. Sci. 53: 2221-2227.

[25]. Reddy, M. and Ghimire, B. (2009). Use of model tree and gene expression programming to predict the suspended sediment load in rivers. J. Intell. Syst. 18: 211-228.

تخمین قیمت مس با استفاده از برنامه‌ریزی بیان ژن

حسام دهقانی

بخش مهندسی معدن، دانشگاه صنعتی همدان، ایران

ارسال ۲۰۱۷/۹/۸، پذیرش ۲۰۱۷/۱۱/۲۹

نویسنده مسئول مکاتبات: dehghani@hut.ac.ir

چکیده:

پیش‌بینی قیمت فلزات در بسیاری از مسائل اقتصادی اهمیت دارد. از سوی دیگر قیمت فلز از جمله متغیرهای حیاتی در مدل‌های مالی ارزیابی درآمد است که تخصیص بهینه از منابع را در پی دارد. با توجه به تغییرات شدید قیمت فلز در سال‌های اخیر، روش‌های برآورد کلاسیک نمی‌توانند به طور صحیح تغییرات را پیش‌بینی کنند. برای حل این مشکل لازم است از الگوریتم‌های هوشمند مصنوعی استفاده شود که توانایی خوبی برای پیش‌بینی نوسانات پدیده‌های مختلف دارند. در پژوهش حاضر، روش برنامه‌نویسی بیان ژن (GEP) برای پیش‌بینی تغییرات قیمت مس مورد استفاده قرار گرفته است. برای درک توانایی این روش، نتایج به دست آمده با نتایج حاصل از دیگر روش‌های کلاسیک تخمین مقایسه شده است. نتایج نشان داد که روش GEP نسبت به روش‌های سری زمانی و رگرسیون چند متغیره از دقت بیشتری برخوردار است.

کلمات کلیدی: قیمت مس، برنامه‌نویسی بیان ژن، تخمین، سری‌های زمانی.