COMPARISON OF NEURAL NETWORKS AND REGRESSION TIME SERIES IN ESTIMATING THE DEVELOPMENT OF THE AFTERNOON PRICE OF GOLD ON THE NEW YORK STOCK EXCHANGE

Veronika MACHOVÁ, Tomáš KRULICKÝ, Jakub HORÁK

Abstract

Gold is a very important commodity in today's global world. Therefore, the price of gold and its development is a fundamental question for many researchers. This paper aims to perform a regression analysis of the development of the afternoon price of gold on the New York Stock exchange using artificial neural networks and linear regression. Data from a period longer than ten years are used. This is a total of 2,578 pieces of data. We use linear regression with the linear, exponential, polynomial, logarithmic, numbers of weighted distances, multiple negative-exponential extermination and spline function. Multilayer perceptron neural networks and neural networks of the radial basis function are generated. A total of 1,000 neural structures are generated, 5 of those with the best characteristics are retained Regarding simple linear regression, the curve obtained via the spline function mirrors the development of the gold price best. However, better results are achieved by all 5 preserved neural networks.

Key words

artificial neural networks, regression time series, prediction, gold price, commodity, future price development

JEL Classification: C22, C45, C53

Introduction

Gold is a chemical element, the atomic number of which is 79 and the mark of which is Au (Latin name Aurum). It is located in the rocks or in the floodplains of many rivers. The color of gold is bright yellow. Gold features include density, softness and shine. Gold is highly ductile and its content is measured in carats. Gold is traded in Troy ounces (Ferry, 2016). Gold has several functions in the world economy as well. It is used in industry and can be converted into jewelry. In the current economic sphere, it works as security against inflation and a safe haven during crises (Ghazali, Lean and Bahari, 2013).

Gold also has other distinctive features. Its supply become accumulated over the years, and its worldwide annual physical output can be as high as 2% of its total supply, so unlike other commodities, its annual output does not necessarily influence its price too much (Reboredo and Ugolini, 2017; Vochozka et al., 2019). In addition, unlike the prices of shares and bonds, the price of gold depends on future supply and demand and is therefore promising (Aye et al., 2015). The conditions of gold market can also affect the balance between the prices of silver and gold where countries are specified by the comparing of current gold price with historical prices. Competitive coefficients appear in particular on the market with high gold prices, suggesting a new long-term balance (Zhu, Peng and You, 2016).

Literature overview

According to Hauptfleisch, Putninš and Lucey (2016), the gold price is influenced by several factors, which are namely the following: changes within the structure and liquidity of markets and macroeconomic announcements. On the base of these findings the authors have discovered that the Futures Market in New York plays the most important role in setting the prices of gold. This center is known as the main gold trading center. Gangopadhyay, Jangir and Sensarma (2016) argue that the factors that affect the gold price include namely inflation, exchange rates, bond prices, market performance, seasonality, revenues, oil prices, and business cycles.

Artificial neural networks, among other tools, can also be used for the purpose of determining, tracking, and predicting the prices of gold. According to Sánchez and Melina (2015), artificial neural networks are currently widely used and can be potentially used in many areas. The analysis of the time series, which this paper deals with, is an area where neural networks are really widely in use.

Neural networks try to capture the behavior of time series and predict individual data points in the best possible way (Sheikhan et al., 2013). In order to predict outputs of the systems with high precision and velocity, it is possible to design the models of time series that are based on neural network processes.

Hu and Hwang (2002) argue that for proper work with the time series, it is essential for neural networks to learn the right way. The effectiveness of the proposed time series and the learning process seems to be a very useful tool for predicting a very complex nonlinear time series (Chen, Yang and Dong, 2006). Neural networks are able to analyze complex patterns very quickly and with high precision and are flexible in their own use (Santin, 2008; Vochozka, 2017). They can be used not only to solve time series but also to understand and generate languages, recognition of texts, etc. (Boguslauskas, 2009). The disadvantage of neural networks is their demand for large sample data because a lot of test observations are needed to create such large data, which is very uncomfortable for users (Stehel, Vrbka and Rowland).

Goal and Methodology

The aim of the paper is to perform a regression analysis of the development of the afternoon gold price on the New York Stock Exchange using artificial neural networks and linear regression. Both methods will be subsequently professionally compared in order to select a more appropriate one to predict the future development of the afternoon gold price.

The data needed for the analysis are obtained from the websites of the New York Stock Exchange or the World Bank (World Bank, 2017). There are specifically used the afternoon gold prices, i.e. London Fix Price PM, which occurred between 3 January 2006 and 15 April 2016. This is a total of 2,578 pieces of data. London Fix Price is the key value for determining the gold reference price. It is also often referred to as London Golden Fix or London Fix. This value is usually announced twice a day, which is in the days when gold is being traded. Morning prices are referred to as London Fix AM and are announced at 10:30. Afternoon prizes announced at 15:00 are referred to as London Fix PM. The pricing has been in place since 1919 in cooperation with the five largest traders on the stock market: Scotland-Mocatta, Barclays Capital, Deutsche Bank, HSBC and Société Générale. The London Fix process itself starts by proposing an opening price, which is close to the spot price, by the chair of the committee. The individual members of the Commission then contact their sales departments and decide who will participate in the selling and purchasing of gold bars at a given price and what quantity it will be. The members may slightly adjust the price so that the supply and demand for gold, the process of which involves the five traders mentioned above, were balanced and there was no overlap between demand and supply. The next step is a determination of London Fix. It usually takes 10-20 minutes, the price is determined in US Dollars (USD), British Pounds (GBP) and Euros (EUR) per troy ounce (Oz is equivalent to 31.1034807 grams). The largest traders trade an estimated 20 tons of gold at the determined price. However, there is no official quantity available. The descriptive characteristics of the data set are given in Table 1.

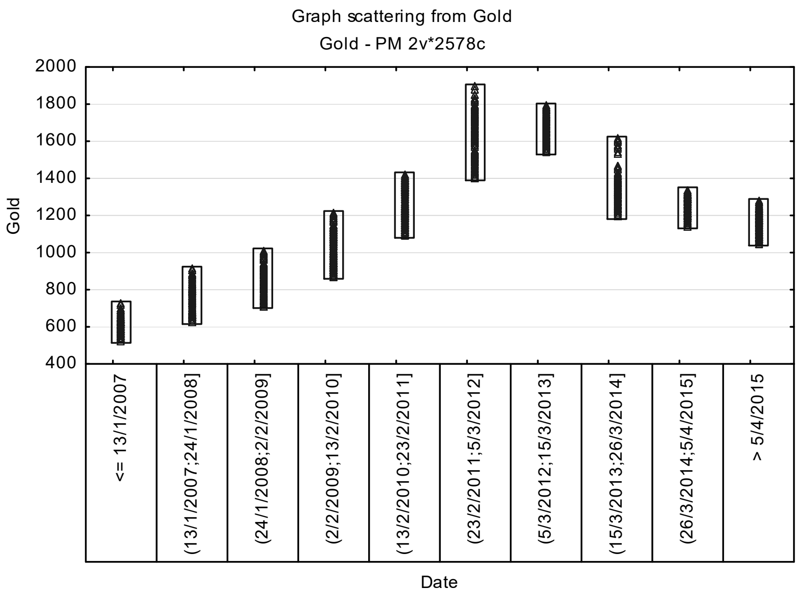
Table 1. Characteristics of the data set

Descriptive Characteristics	Value at USD
Minimal Value	524.75
Maximal Value	1,895.00
Mean Value	1,145.87
Dispersion	119,220.87

Source: Own processing

The difference between the minimum and maximum value was relatively large in the monitored period. This is also due to the world economic crisis and instability in the financial world. For these reasons, the price of gold has grown significantly. It is well known that the price of gold grows in economic crises, instability, and negative announcements. The average price of gold was nearly USD 1,146 per troy ounce in the monitored period. The price development over time is interesting, of course. Figure 1 therefore provides a view of the dispersion of values in the individual periods of the observed time period.

Fig. 1. Graph of the dispersion of gold price (London Fix PM)



Source: Own processing

The data processing will be done using DELL's version of Statistica, version 12. In the first step, linear regression will be performed, while the following functions will be used:

- linear,
- exponential,
- polynomial,
- logarithmic,
- numbers of weighted distances,
- multiple negative-exponential extermination,
- spline.

The correlation coefficient will be calculated. It corresponds to the dependence of the gold price on time. Furthermore, the level of significance will be at 0.95.

The regression will follow with the help of artificial neural networks. Multilayer Perceptron Networks (MLP) and radial basic functions (RBF) will be generated. The independent variable will be time, the dependent variable will be the price of gold. The time series will be divided into three sets – training (which includes 70% of input data), test (15% of input data) and validation (15% of input data). Based on the training set of data, neural structures will be generated, the test and validation groups will be available to verify the reliability of the identified neural network, or discovered model. The delay of the time series will be set at the value of 1. A total of 1,000 neural structures will be generated, 5 of those with the best characteristics will be retained. The hidden layer of multilayer perceptron networks will contain at least two neurons, but not more than 50. In case of a radial base function, at least 21 neurons, at most 30, will be hidden in the hidden layer. These distribution functions will be in the hidden and output layers:

- identity,
- logistics,
- atanh,
- exponential,
- sinus.

Other settings will be left default – ANS, an automated neural network. In conclusion, the results of linear regression and regression will be compared using neural networks. The comparison will not take place in the form of residue analysis (minimum values, maximum values, dispersion of residues, etc.), but at expert level and experience of the evaluator, economist.

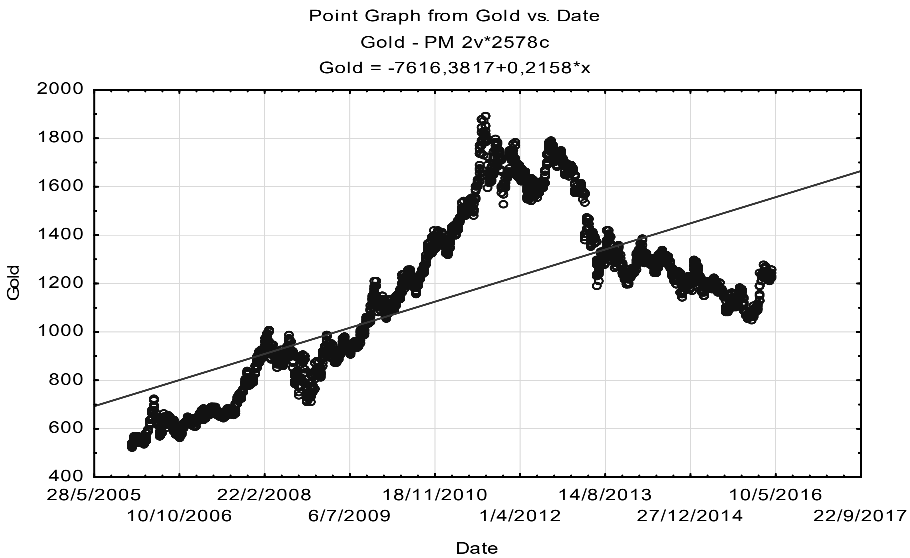
Results and Discussion

As it has been suggested several times, the results will include a part devoted to linear regression and a part dealing with regression using neural networks.

1 Linear regression

The correlation coefficient is set at 0.6781, which means a significant statistical dependence of gold on the development over time. Figure 2 is a scatter plot where the points are fitted with a regression curve, in this case linear. The line parameters are shown in the figure.

Fig. 2. Scatter plot of gold prices fitted with regression curve – linear function



Source: Own processing

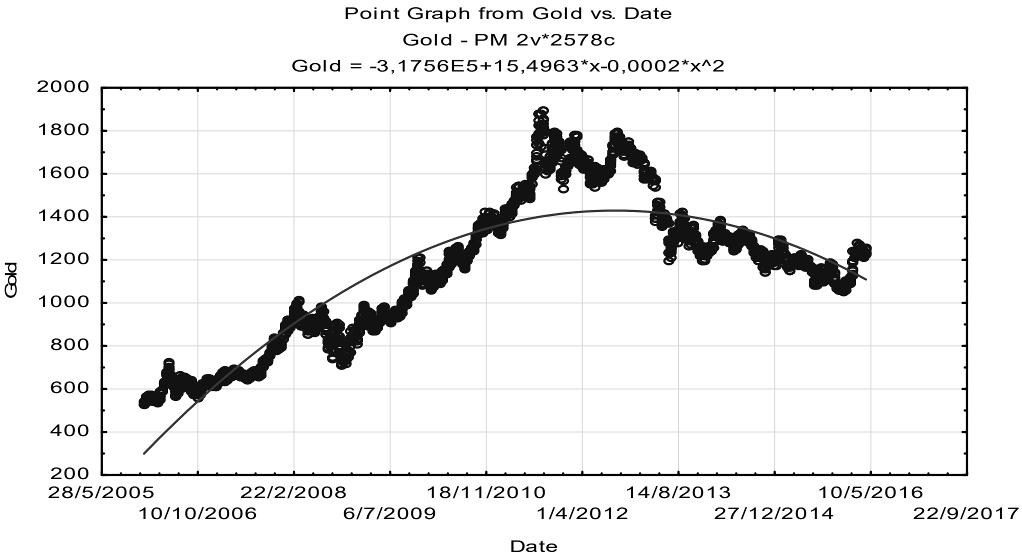
Here, the logistic function has a following form:

$$y = a + b(x) \qquad (1)$$

The actual development of the price of gold is represented by blue points, while the red curve is a regression curve, which is the linear function in this case. The figure shows that the linear function is not able to capture the development of the price of gold at all. Therefore, for forecasting the prices of gold, the logistic function is totally unsuitable.

Figure 3 refers to the interleaving of a scatter plot by polynomial function.

Fig. 3. Scatter plot fitted with regression curve – polynomial function



Source: Own processing

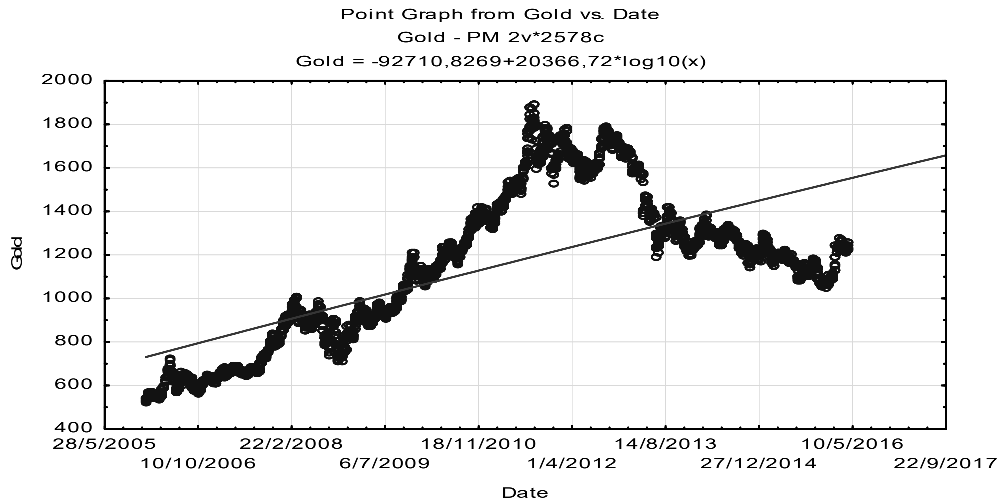
In linear regression, polynomial function represents fitting the values entered by a polynomial, where the coefficients of the polynomial sought are calculated using the method of least squares so that the sum of the square power of the original values deviations from the polynomial obtained are minimal:

$$y = a + b(x) + \cdots + n \qquad (2)$$

The figure clearly shows that in the case of regression analysis, polynomial function is not able to follow the development of the price of gold in the monitored period.

Figure 4 shows a scatter plot fitted with the logarithmic function.

Fig. 4. Scatter plot fitted with regression curve – logarithmic function



Source: Own processing

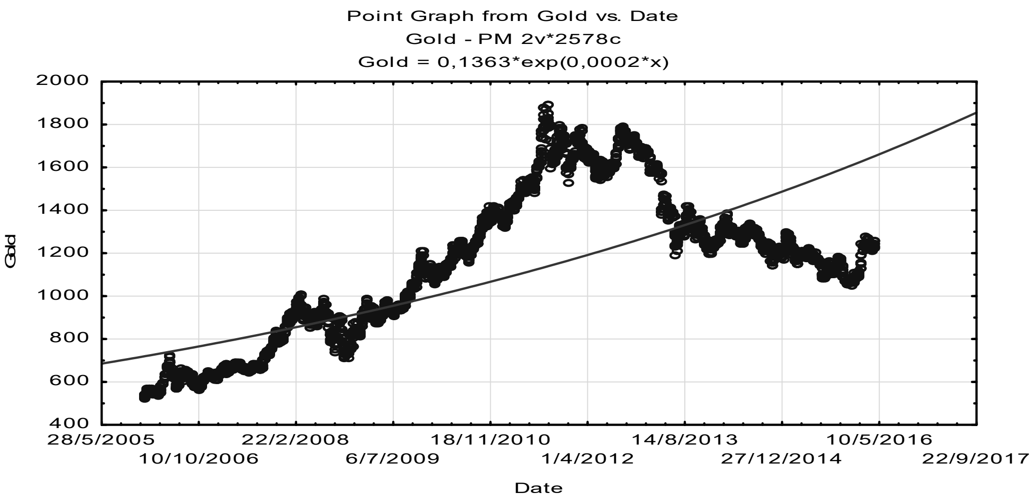
In this case, it is a linear regression, where the dataset was fitted by logarithmic function:

$$y = a + b * \ln(x) \qquad (3)$$

The course of the curve fitted with a logarithmic function is very similar to the course of the linear function. It is therefore a completely inadequate expression of the price of gold development. This function is thus insufficient for forecasting the future development of the price of gold as well.

Next figure 5 shows the scatter plot of the afternoon price of gold fitted with an exponential function.

Fig. 5. Scatter plot fitted with regression curve – an exponential function



Source: Own processing

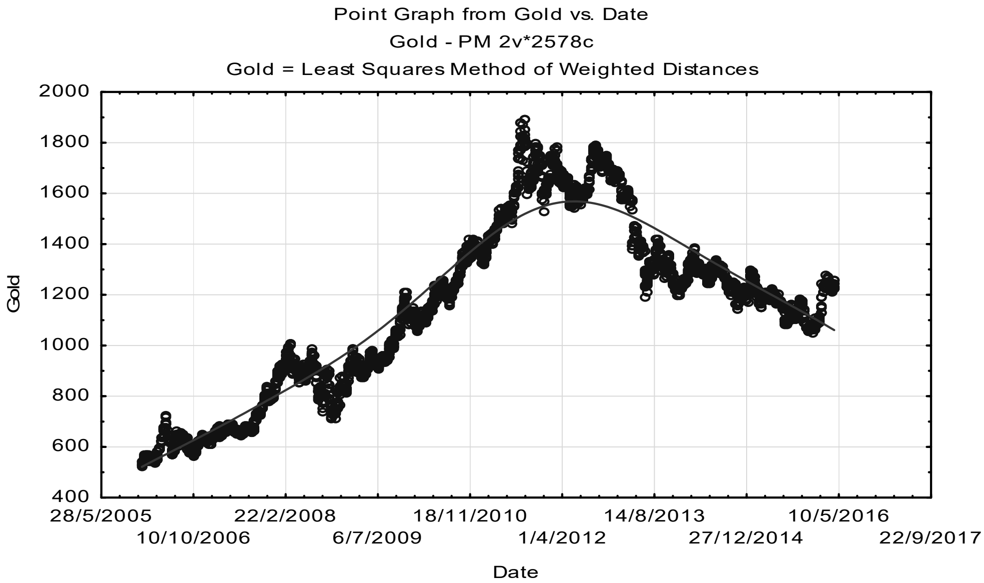
It is a linear regression where the dataset was fitted with the exponential function in the following form:

$$y = ab^x \tag{4}$$

However, it is not linear in terms of its parameters but by a suitable transformation, it can be converted into a form that it is linear in its parameters. Even the exponential function cannot follow the development of the price of gold for the given period of time; this function is thus unsuitable for forecasting the price of gold, and can only be used as a starting point for an iteration algorithm.

Figure 6 offers a London Fix Price PM dotted by a function obtained by the least squares method of weighted distances.

Fig. 6. Scatter plot fitted with a regression curve – a function of the smallest squares of weighted distances



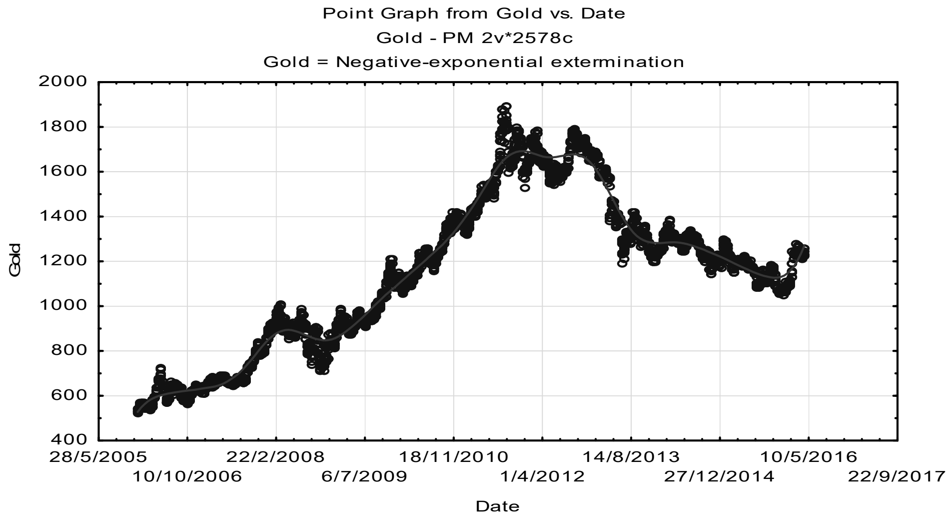
Source: Own processing

The method of least squares provides sufficient estimations of parameters only when all conditions on the data and regression model are met. If these requirements are not met, the results of the method of least squares lose their properties. The method of least squares is generally a mathematical-statistical method for the approximation of solving systems of equations where there are more equations than unknown variables. Least squares mean that the result shall minimize the sum of deviation squares to each equation.

The figure shows that the function of the method of least squares of weighted distances only roughly follows the actual development of the stock price. However, it shows the best results from the curves observed so far. Still, it is not applicable in practice.

The scatter plot fitted with the function obtained by the method of least squares by negative-exponential smoothing is shown in Figure 7.

Fig. 7. A scatter plot fitted with a regression curve – a function of the method of least squares by a negative-exponential smoothing

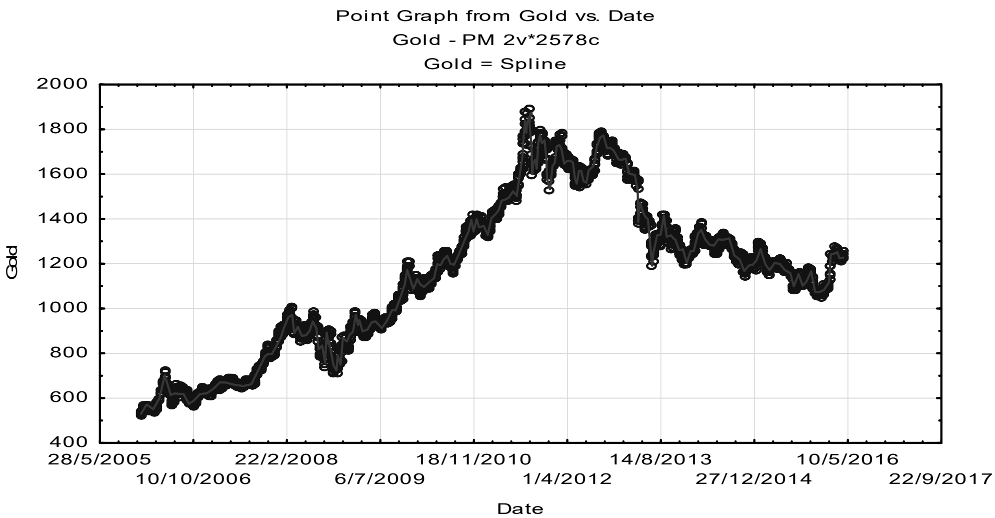


Source: Own processing

The method of least squares through negative exponential smoothing is very similar to the previous method, as it works on a similar principle. There is a slight difference in the compilation of the curve on the basis on the mathematical derivation. The figure shows that this function is able to follow the actual development of the price of gold in the given period of time very roughly, but it is not able to capture the local minimum and maximum of the time series. In practice, the application of the method of least squares through negative exponential smoothing in regression analysis is very limited.

The scatter plot fitted with the spline function is the subject of figure 8.

Fig. 8. Scatter plot fitted with a regression curve – spline



Source: Own processing

The spline function is generally used for fitting almost any regression curves with the measured data with a one-dimensional independent variable x and one-dimensional random dependent variable y in the following form:

$$S_1(x) = \frac{y_i + 1(x - x_i)}{h_i} + y_i\left(1 - \frac{(x - x_i)}{h_i}\right), x \in \langle x_i, x_i + 1\rangle, i = 0 \quad (5)$$

where the linear function spline is the function $S_1(x)$, which is continuous at the interval of $\langle x_0, x_n\rangle$ and at each interval $\langle x_i, x_i + 1\rangle, i = 0, \ldots, n$ $S_1(x)$ is a first degree polynomial. As evident from the figure, it is the best model from the linear regression analysis carried out. The spline function is able to follow the development of the price of gold in the monitored period relatively precisely, and is able to capture the local minimum and maximum of the time series rough. It appears to be definitely the best model from the linear regression models used.

It has already been stated above that the correlation coefficient indicates a significant statistical dependence of the target variable on the development over time. If we want to evaluate the results only by comparing the development of the London Fix Price PM and the shape of the regression curve (assuming a simple linear regression), we can safely say that the curve obtained by the spline is closest to the development. A curve obtained by the method of least squares by a negative-exponential smoothing is quite appropriate, followed by a curve obtained yet again by the method of least squares, this time by weighted distances. All of these curves quite reliably copy the basic development of the gold price. The curve obtained by the spline function even tracks not only the global extremes of London Fix Price PM but also the local extremes of this development as well. Optically, this function appears to be effective, given the possible prediction of the gold price development.

2 Neural structures

Based on the methodology, 1,000 neural structures were generated, 5 of which were retained. These networks have the best parameters. Their overview is shown in Table 2.

Table 2. Overview of the retained neural networks

Net work name	Tr aining perf.	Te sting perf.	Va lid. perf.	Tr ain. error	Te st. error	Va lid. error	Train. algorithm	E rror function	Activ ation of hidden layer	Out put activation layer
MLP 1-6-1	0.9 99117	0.9 99279	0.9 98911	10 6.8757	79. 99031	12 8.8157	BFGS (Quasi-Newton) 20	S um. of square	Expo nential	Ide ntity
MLP 1-8-1	0.9 99112	0.9 99281	0.9 98912	10 8.0174	79. 21815	12 9.9734	BFGS (Quasi-Newton) 11	S um. of square	Atan h	Ide ntity
MLP 1-7-1	0.9 99116	0.9 99282	0.9 98911	10 7.0119	79. 77784	12 8.8021	BFGS (Quasi-Newton) 20	S um. of square	Expo nential	Ide ntity
MLP 1-6-1	0.9 99113	0.9 99282	0.9 98912	10 7.3197	79. 70958	12 8.9285	BFGS (Quasi-Newton) 13	S um. of square	Atan h	Ide ntity
MLP 1-8-1	0.9 99109	0.9 99281	0.9 98912	10 7.8812	79. 62574	12 9.4264	BFGS (Quasi-Newton) 14	S um. of square	Atan h	Ide ntity

Source: Own processing

The table above shows that these are only multi-layer perceptron networks with one hidden layer. There is a single variable in the input layer, while the hidden layer contains neural structures from 6 to 8 neurons. The output layer contains logically only one neuron and one output variable – London Fix Price PM. The Quasi-Newton training algorithm has been applied to all preserved networks. Mutually, the neural networks differ in the type of activation functions used in the hidden layer of the neurons. If we focus on training, testing, and validation performance, we generally look for a network that will have the performance across all data sets, (note the partitioning of data into sets was done randomly), ideally the same, with the error as small as possible. This performance of individual sets of data is given in the form of a correlation coefficient. The values of individual data sets according to specific neural structures are shown in Table 3.

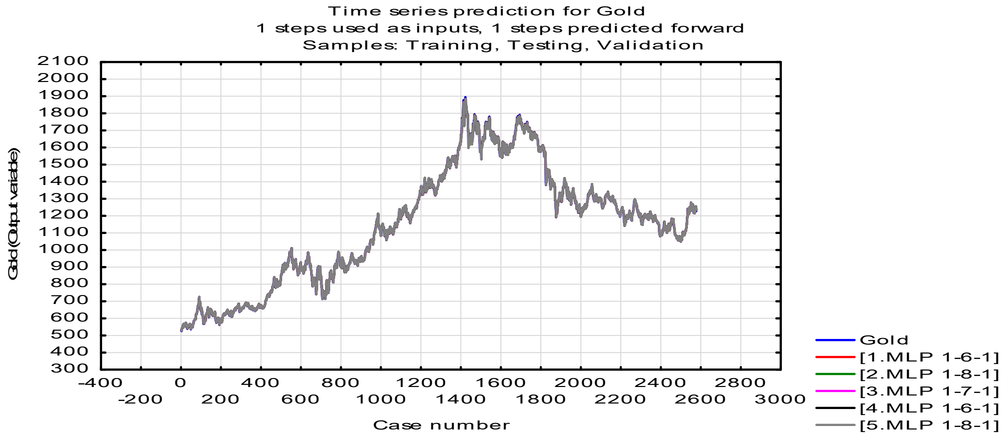
Table 3. Correlation coefficient

	Palladium Training	Palladium Testing	Palladium Validation
1.MLP 1-12-1	0.999117	0.999279	0.998911
2.MLP 1-4-1	0.999112	0.999281	0.998912
3.MLP 1-7-1	0.999116	0.999282	0.998911
4.MLP 1-3-1	0.999113	0.999282	0.998912
5.MLP 1-6-1	0.999109	0.999281	0.998912

Source: Own processing

Consequently, the performance of all conserved neural networks is approximately identical. Slight differences do not affect network performance at all. Figure 9 shows a conjugal graph showing the true development of London Fixed Price PM (note in the figure it is indicated as 'Gold') and the development of predictions using individual generated networks (note these are indicated by the order number given in Table 2 and the number of neurons in each layer). On closer examination of the image, it is clear that all generated neural networks predict the afternoon gold price development very similarly, although deviations can be registered at first glance (e.g. for networks 3 and 4 for 600 observations). However, it is not the similarity of the forecasts of individual networks, but the similarity, i.e. the degree of consistency, with the real development of the gold price that really matters. Even in this respect, preserved neural structures appear to be very interesting at first glance, as they respect the global extremes of the curve assessing the development of the afternoon gold price on the New York Stock Exchange. However, they also tend to register the local extremes of this curve.

Fig. 9. Conjunctural chart – Gold price trend predicted by neural networks compared to actual price in the reference period



Source: Own processing

Conclusion

Generally, every prediction is given by a certain degree of probability with which it is to be filled. At a time when we predict the future development of any variable, we try to predict the future development of this variable based on previous years' data. Even though most of the factors influencing the target quantity can be included in the model, there is always a certain simplification of reality. So, we always work with a certain degree of probability that the predicted scenario will be fulfilled. The article used linear regression and regression using neural networks to meet the target. Even in these two cases, however, there is a relatively simple simplification. The author works only with two variables – input (time) and output (gold price). It is more than clear that there is a disregard for other input quantities, which undoubtedly influence the final price of gold. These are, for example, the development of the national economy, the political situation of the state, the legal environment, market barriers etc. In spite of, or precisely because there are a number of factors influencing the price of gold, we have to think whether working with time series does not simplify the development of the target variable too. Or, on the contrary, the other variables are so insignificant that the input quantity (in this case time) and the output value (London Fix Price PM) are sufficient. To predict the emergence of extraordinary situations and their influence on the price of gold is almost impossible. In the short term it is possible, but in the long run certainly not. Simplification and the creation of a relatively simple model are therefore appropriate and therefore the result is useful.

The commodity price – gold – can be determined on the basis of statistical methods, causal methods and intuitive methods. This paper was devoted to the comparison of individual statistical methods. However, we have only provided a possible framework for the development of the afternoon gold price. It is very important to work with information on the possible future development of both the economic and the political or legal environment. If it is possible to predict its future development, it can then be reflected in the gold price. At the same time, however, the personality of the evaluator, an economist who, on the basis of expert knowledge and experience, corrects the price determined by the statistical methods and specified on the basis of causal links.

The aim of the paper was to perform a regression analysis of the development of the afternoon gold price on the New York Stock Exchange using artificial neural networks and linear regression. Then the intention was to compare the two methods more appropriately in order to predict the future development of the afternoon gold price. From the results described, it appears that the curve obtained by the spline function appeared to be optically best from the linear regression. Artificial neural networks all proved to be useful in practice. If we look at performance from a correlation coefficient point of view, only neural networks are left to use, with virtually no difference in terms of practical use. Certainly, it would be interesting to analyse residues that would certainly help to determine the best of preserved neural structures. This, however, was not the aim of the contribution, and may be the subject of further research.

References

Aye, G., Gupta, R., Hammoudeh, S., Kim, W. J. (2015). Forecasting the price of gold using dynamic model averaging. *International Review of Financial Analysis*, 41, 257-266.

Boguslauskas, V., Mileris, R. (2009). Estimation of Credit Risk by Artificial Neural Networks Model's. *Engineering Economics,* 64 (4), 7-14.

Chen, Y., Yang, B., Dong, J. (2006). Time-series prediction using a local linear wavelet neural network. *Neurocomputing,* 69 (4-6), 449-465.

Ferry, E. (2016). Gold prices as material-social actors: The case of the London Gold Fix. *The Extractive Industries and Society,* 3 (1), 82-85.

Gangopadhyay, K., Jangir, A., Sensarma, R. (2016). Forecasting the price of gold: An error correction approach. *IIMB Management Review,* 28 (1), 6-12.

Ghazali, H., Bahari, Z. (2013). Is gold a hedge or a safe haven? An empirical evidence of gold and stocks in Malaysia. *International Journal of Business and Society,* 14 (3), 428-443.

Hauptfleisch, M., Putniņš, T. J., Lucey, B. (2016). Who Sets the Price of Gold? London or New York. *Journal of Futures Markets,* 36 (6), 564-586.

Hu, Y. H., Hwang, J. N. (2002). *Handbook of neural network signal processing*. Boca Raton: CRC Press. ISBN 978-084-9323-591.

Reboredo, J. C., Ugolini, A. (2017). Quantile causality between gold commodity and gold stock prices. *Resources Policy*, 53, 56-63.

Santin, D. (2008). On the approximation of production functions: a comparison of artificial neural networks frontiers and efficiency techniques. *Applied Economics Letters,* 15 (8), 597-600.

Sánchez, D., Melin, P. (2015). Modular Neural Networks for Time Series Prediction Using Type-1 Fuzzy Logic Integration. *Studies in Computational Intelligence*, 141-145.

Sheikhan, M., Mohammadi, N., Tawfik, H., Ben-Jacob, E. (2013). Time series prediction using PSO-optimized neural network and hybrid feature selection algorithm for IEEE load data. *Neural Computing and Applications,* 23 (3-4), 1185-1194.

Stehel, V., Vrbka, J., Rowland, Z. (2016). Using neural networks for determining creditworthiness for the purpose of providing bank loan on the example of construction companies in South Region of Czech Republic. *Ekonomicko-manažerské spektrum,* 2016 (2), 62-73.

Vochozka, M., Stehel, V., Rowland, Z. (2019). Determining development of business value over time with the identification of factors. *Ad Alta: Journal of Interdisciplinary Research*, 9 (2), 358-363.

Vochozka, M. (2017). Comparison of neural networks and regression time series in estimating the development of the afternoon price of palladium on the New York Stock Exchange. *Trends Economics and Management*, 30 (3), 73-83.

World Bank. (2017). *The World Bank*. [on-line] [cit.: 2020-01-20]. Retrieved from http://www.worldbank.org/.

Zhu, H., Peng, C., You, W. (2016) Quantile behaviour of cointegration between silver and gold prices. *Finance Research Letters,* 19 (C), 119-125.

Contact

Ing. Veronika Machová, MBA
School of Expertness and Valuation, The Institute of Technology and Business in České Budějovice
Okružní 517/10, 370 01 České Budějovice, Czech Republic
machova@mail.vstecb.cz

Ing. Tomáš Krulický, BBA
Faculty of Operation and Economics of Transport and Communications, University of Žilina
Univerzitná 8215/1, 010 26 Žilina, Slovakia
krulicky@mail.vstecb.cz

Ing. Jakub Horák
Faculty of Operation and Economics of Transport and Communications, University of Žilina
Univerzitná 8215/1, 010 26 Žilina, Slovakia
horak@mail.vstecb.cz