

Daily Forecasting for Antam's Certified Gold Bullion Prices in 2018-2020 using Polynomial Regression and Double Exponential Smoothing

**Tresna Maulana Fahrudin¹, Prismahardi Aji Riyantoko², Kartika Maulida
Hindrayani³, I Gede Susrama Mas Diyasa⁴**

Pembangunan Nasional "Veteran" University of East Java^{1,2,3,4}
Raya Rungkut Madya Street, Gunung Anyar Sub-district, Surabaya, East Java 60294,
Indonesia

Correspondence Email: tresna.maulana.ds@upnjatim.ac.id

ABSTRACT

Gold investment is currently a trend in society, especially the millennial generation. Gold investment for the younger generation is an advantage for the future. Gold bullion is often used as a promising investment, on other hand, the digital gold is available which it is stored online on the gold trading platform. However, any investment certainly has risks, and the price of gold bullion fluctuates from day to day. People who invest in gold hopes to benefit from the initial purchase price even if they must wait up to five years. The problem is how they can notice the best time to sell and buy gold. Therefore, this research proposes a forecasting approach based on time series data and the selling of gold bullion prices per gram in Indonesia. The experiment reported that Holt's double exponential smoothing provided better forecasting performance than polynomial regression. Holt's double exponential smoothing reached the minimum of Mean Absolute Percentage Error (MAPE) 0.056% in the training set, 0.047% in one-step testing, and 0.898% in multi-step testing.

Keywords: Double Exponential Smoothing, Forecasting, Gold Bullion Prices, Polynomial Regression, Time Series

INTRODUCTION

The era of advances in information technology and digital transformation makes it easier for people to do investments, one of which is gold investment. Indonesia as a country with potential natural resources and the largest gold mine in the world can support economic turnover through business and gold investment. Indonesia has potential gold mining in various locations located in Papua (Mulyana & Djajadiningrat, 2013), Sumbawa, East Kalimantan, and Central Kalimantan (Singawinata, 2005). Global mining companies even given Indonesia a high rating as the holder of a significant role in the gold, tin, coal, copper, nickel and bauxite mining industry. Although there are some limited investments in mining and this is related to greenfield projects, the mining industry is still doing well in Indonesia (PwC, 2018).

Various products that can be produced from processed gold mining include gold bullion, gold coins, jewelry, and accessories. The quality of gold mining in Indonesia is even known to have good quality and grade (Basov, 2015). Nowadays, there is a change in the way of collecting gold jewelry and accessories, which is favored by generation X over the age of 40 in Indonesia, while now generations Y and Z prefer to collect various brands

and sizes of gold bullion. Millennials in other countries are also aware of the importance of investing in gold for the long term (Karanam & Shenbagavalli, 2019; Street, 2019). Gold bullion is often used as a promising investment, even the digital gold is available which does not have a physical form and stored online on the gold trading platform (Pusca, Tureac, Andronic, & Filip, 2012). Gold investing for beginners is easier and digital gold savings can be exchanged for physical gold.

Gold investment for younger generation is an advantage for the future. Several brand choices of gold bullion products are well known to the Indonesian people with sizes of 0.5 grams to 100 grams, such as ANTAM, UBS, and Lotus Archi. There are newcomer products that are quite competitive by carrying small gold bullion products such as MiniGold with the smallest size of 0.01 grams. This has become a new investment model for millennials besides investing in property and stocks. However, any investment certainly has risks, such as the price of gold bullion fluctuates from day to day.

People who invest in gold hopes to benefit from the initial purchase price even if they have to wait up to five years. The problem is how to recognize the fluctuations in the future so that people can know the right time to sell and buy gold. Several researchers have proposed various approaches and methods to obtain the best prediction model. Mombeini & Yazdani-Chamzini (2015) proposed forecasting gold price changes in 1990-2008 using an Artificial Neural Network (ANN) compared to traditional statistical methods such as ARIMA. The research showed that ANN was better than ARIMA based on performance measurements using Mean Absolute Error (MAE), Coefficient of Determination (R²), and Root Mean Square Error (RMSE).

Aye, Gupta, Hammoudeh, and Kim (2015) developed several prediction models involving six global factors (interest rate, nominal, business cycle, exchange rate, commodity, and stock price), the U.S. economic policy uncertainty index, and the Kansas City Fed's financial stress index. The research showed that Dynamic Model Averaging and Dynamic Model Selection were better than linear models such as Bayesian Model Averaging based on the Mean Squared Forecast Error (MSFE). All variables represented strong predictive at the time series. Another research by Wang, Sheng and Zhang (2019) developed a prediction model involving supply and demand factors, speculation factors, and financial factors. The research showed that Structural Vector Autoregression (SVAR) was able to identify the factors which affected the international futures of gold prices in China. Speculation and financial factors had a significant effect on fluctuations in the gold prices, people consider gold as financial wealth. Meanwhile, the fundamental role was determined by supply and demand factors.

Therefore, this research proposed a forecasting approach based on time series data and the selling of gold bullion prices per gram in Indonesia using polynomial regression and double exponential smoothing methods. The two methods were compared to obtain the best model with the small Mean Absolute Percentage Error (MAPE), and to utilize the best model as a prediction model for future gold bullion prices.

RESEARCH METHOD

This section discusses the research system design of daily forecasting for gold bullion prices, including population and samples. In addition, it discusses the theory of polynomial regression, double exponential smoothing, two methods of regression analysis for building the predictive models. The model will be evaluated by mean

absolute percentage error as the forecasting performance measures. The evaluation scenario is divided into two stages, namely the training and testing set. The research system design was illustrated in Figure 1.

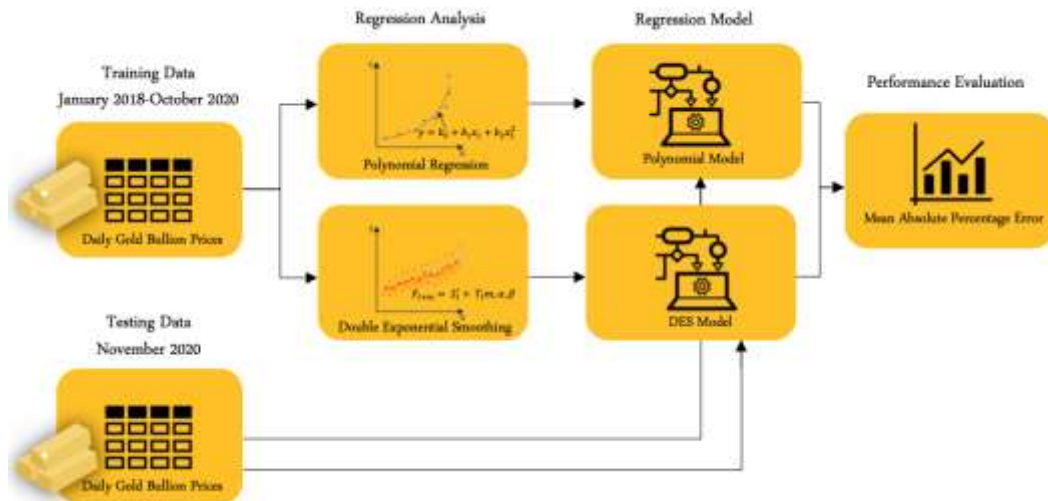


Figure 1. System Design of Daily Forecasting for Gold Bullion Prices

Population and Samples

The dataset was obtained from a company, Logam Mulia Ltd, a seller of gold and silver bullion, dinar, and dirham coins. However, the research focuses on the dataset of ANTAM LM (product brand) gold bullion prices per gram which can be accessed at <https://www.logammulia.com/id/purchase/gold>. Training data were taken from January 2018 to October 2020, while data testing was conducted on November 2020. Figure 2 showed the official website of Logam Mulia Ltd which present the daily gold bullion prices per gram. For example, the gold bullion prices on November 23, 2020 per gram is IDR 977,000. That means the X-axis is the date (time series) and Y-axis is gold bullion prices.



Figure 2. Official Website of Logam Mulia Ltd for Collecting Daily Gold Bullion Prices

Polynomial Regression

Building a time series-based regression model with y target must match with the pattern of actual data. There are many options for regression analysis models, including simple linear regression and multiple linear regression. Simple linear regression is applied to find a relationship between one dependent variable and one independent variable, while multiple linear regression is applied to find the relationship between one dependent variable and many independent variables (Uyanık & Güler, 2013). However, there is a different case, if the form of an unfit model which only uses one dependent variable with one independent variable. Therefore, the solution is using polynomial regression (Aslam, Heijden, & Collins, 2019).

To understand the difference formula between simple linear regression, multiple linear regression and polynomial regression as follows.

The formula of simple linear regression follows equation (1).

$$y = a_0 + a_1 * x_1 \quad (1)$$

Where y is variable that will be predicted, a_0 is a constant number, a_1 is coefficient of trend line, and x_1 is time indices (Djatmiati, Alfath, & Fahrudin, 2019).

The formula of multiple linear regression follows equation (2).

$$y = a_0 + a_1 * x_1 + a_2 * x_2 + \dots + a_n * x_n \quad (2)$$

Where the multiple linear regression basically is similar with linear regression yet with many constant numbers and many coefficients of the trend line (Uyanık & Güler, 2013).

While the formula of polynomial regression follows equation (3).

$$y = a_0 + a_1 * x_1 + a_2 * x_1^2 + \dots + a_n * x_1^n \quad (3)$$

Where polynomial regression is a regression which the function is quadratic, the value of the independent variable is the power of 1, the power of 2, and the power of n (Aslam, Heijden, & Collins, 2019).

Figure 3 illustrated the difference between simple linear regression and polynomial regression.

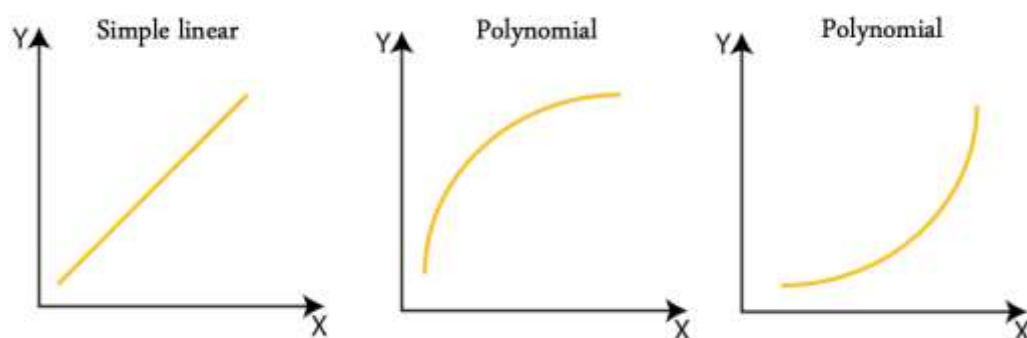


Figure 3. The Difference Between Simple Linear and Polynomial

Polynomial in some literature is referred to as polynomial linear regression. It belongs to a linear function, and not a group of non-linear (Tyrrell, 1983). The definition of linear is a linear combination (chain) relationship between the dependent variable and the coefficient of the independent variable which is consistent. While non-linear is a function

which the coefficient cannot stand alone, for example a_1 requires a_2 (Peddada & Haseman, 2005). The formula non-linear function follows equation (4).

$$y = a_0 + \frac{a_1 * x_1}{a_2 * x_2 + x_3} \quad (4)$$

Double Exponential Smoothing

Double exponential smoothing (DES) has two types, Brown's DES and Holt's DES. Both to solve the drawback of single exponential smoothing, DES can tackle the problems when facing the difficult to forecast the actual data in trend pattern and random variation. Brown's DES makes the trend and slope become smoother using different smoothing constant directly, while Holt's DES adjust the trend and slope rates flexibly using two smoothing constants α and β (Nazim & Afthanorhan, 2014). Basically, Holt's DES is like Brown's DES (Hyndman, Koehler, Ord, & Snyder, 2008), however Holt's DES uses two components. They are exponentially smoothed series following equation (5), and trend estimation follow equation (6).

$$S'_t = \alpha X_t + (1 - \alpha) (S'_{t-1} + T_{t-1}) \quad (5)$$

The S'_t not only calculate the exponential smoothing based on the current observation X_t and the previous exponentially smoothed series S'_{t-1} , but also calculate the previous trend estimation T_{t-1} . To use equation (5) and (6), S'_{t-1} and T_{t-1} must be calculated previously. These values are not available when $t = 1$. Therefore, it must be determined at the beginning of the period where S'_{t-1} and T_{t-1} are equal to X_1 .

$$T_t = \beta (S'_t - S'_{t-1}) + (1 - \beta) T_{t-1} \quad (6)$$

Holt's DES makes the model more flexible to adjust the trend and slope rates, where $0 < \alpha < 1$ and $0 < \beta < 1$ (Bose, Dey, Roy, & Sarddar, 2019).

$$F_{t+m} = S'_t + T_t m \quad (7)$$

To forecast the next period t , it follows the equation (7), where F_{t+m} , m is the period number of a step ahead. The next forecast is the exponentially smoothed series S'_t is added to the product of the trend estimation T_t and the period number of a step ahead m .

Mean Absolute Percentage Error (MAPE)

MAPE is a forecasting performance measures used to calculate the percentage error of the actual (observation) data with the forecast result. MAPE uses the absolute error of each period obtained from the difference between the actual data and the forecast results, then divided by the actual data (Kim & Kim, 2016). The result is the mean of absolute percentage. The formula of MAPE follows the equation (8).

$$MAPE = \left(\frac{100\%}{n} \right) \sum_{t=1}^n \frac{|X_t - Ft|}{X_t} \quad (8)$$

The X_t is actual data in period t , Ft is forecast result in period t , and n is total data.

RESULTS AND DISCUSSION

This section discusses the performance evaluation of forecasting model using Polynomial Regression and Holt's Double Exponential Smoothing. The experiment were divided into two scenarios, training set and testing set. Training set used the gold bullion prices from January 2018 to October 2020, while testing set used the gold bullion prices in November 2020. The performance evaluation measures based on Mean Absolute Percentage Error (MAPE) would have chosen either the best model of regression analysis using Polynomial Regression or Holt's Double Exponential Smoothing.

ANTAM's Certified Gold Bullion Prices

Training data were collected from the official website of Logam Mulia Ltd yet focusing on ANTAM's Certified Gold Bullion Prices. Figure 4 and 5 showed the scatter chart of training data (red dots) taken in January 2018 to October 2020. The total data were 890 days, where the lowest gold bullion price reached IDR 642,000 and the highest reached IDR 1,065,000. In the chart, it can be seen that the gold bullion prices on day-0 to 400 (January 2, 2018 - May 2, 2018) was stable, to hover in the range IDR 600,000-IDR 700,000. The price of gold bullion increased on June 20, 2019 (day-432) the price of gold reached up to IDR 700,000. February 21, 2020 (day-656) reached up to IDR 800,000, and March 23, 2020 (day-689) reached up to 900,000, and July 28, 2020 (day-794) reached up to IDR 1,000,000.

Performance Evaluation of Forecasting Model using Polynomial Regression

First scenario was using training set, the experiment involved the gold bullion prices in January 2018 to October 2020. Polynomial regression was applied as the model of regression analysis, which the number of degrees set = 6 in the polynomial regression. Figure 4 showed the intersection point between actual data and forecast results. Although the intersection points look like they were close together, the forecast line was still not precise enough to follow the actual data pattern.



Figure 4. The Forecasting Result of ANTAM's Certified Gold Bullion Prices from January 2018 to October 2020 using Polynomial Regression in Training Set

Table 1 showed the MAPE report of ANTAM's certified gold bullion prices using Polynomial Regression using degree 1-6. In the training set, the higher degree set into the polynomial regression, the smaller MAPE. Minimum MAPE was obtained using degree 6 in training set which reached 2.236%. Likewise, in the testing set, although there was a fluctuation in degree 3, polynomial regression using degree 6 given the best performance in the testing set for one-step and multi-step reached minimum MAPE 2.206% and 1.165% respectively.

Table 1. MAPE of ANTAM's Certified Gold Bullion Prices using Polynomial Regression

Degree	Training set	Testing set	
		One-step	Multi-step
1	6.166%	2.463%	6.107%
2	2.509%	2.569%	8.998%
3	2.396%	2.614%	10.006%
4	2.206%	2.276%	6.672%
5	2.277%	2.306%	4.236%
6	2.236%	2.206%	1.165%

Performance Evaluation of Forecasting Model using Double Exponential Smoothing
 Polynomial regression needs to be adjusted to the right degree to obtain the smallest MAPE. On the other hand, Holt's double exponential smoothing needs to determine the optimal α and β values. Figure 5 showed the intersection point between actual data and forecast results. The intersection points look like they were close together, the forecast line was very precise following the actual data pattern. It is better than using polynomial regression.

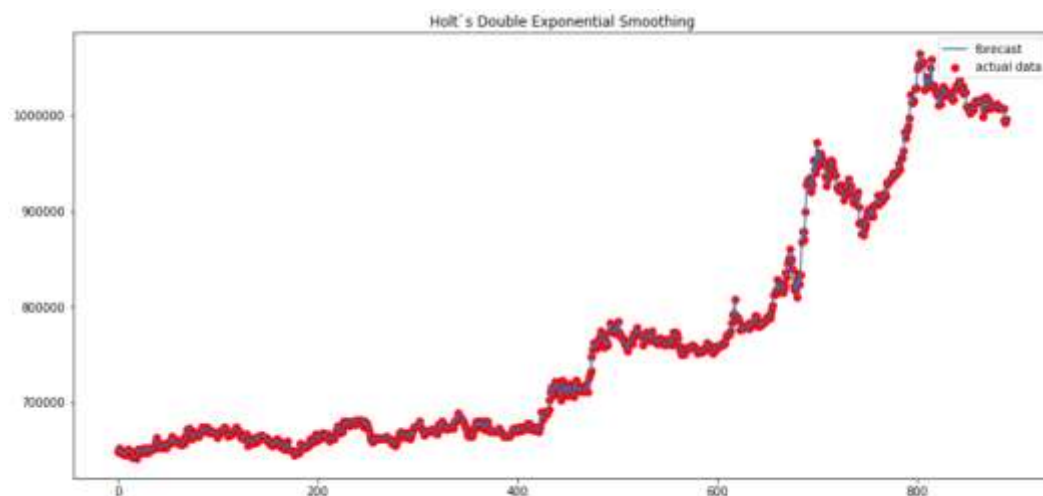


Figure 5. The Forecasting Result of ANTAM's Certified Gold Bullion Prices from January 2018 to October 2020 using Holt's Double Exponential Smoothing in Training Set

Table 2 showed the MAPE report of ANTAM's certified gold bullion prices using Holt's double exponential smoothing using degree α and β . Minimum MAPE was obtained in training set which reached 0.056%, while Holt's double exponential smoothing given the best performance in the testing set for one-step and multi-step reached minimum MAPE 0.047% and 0.898% respectively.

Table 2. MAPE of ANTAM's Certified Gold Bullion Prices using Double Exponential Smoothing

Optimal Parameter	Training Set	Testing Set	
		One-step	Multi-step
$\alpha = 0.9, \beta = 0.1$	0.056%	0.047%	0.898%

CONCLUSIONS

To predict the gold bullion prices, at least 2 axes are needed, namely the X-axis as the date and the Y-axis as the gold prices. Likewise, this research involved actual data on gold bullion prices from January 2018 to October 2020 as training data, while the testing data involved data in November 2020 using one-step and multi-step testing. The experiment reported that Holt's double exponential smoothing provided the better performance than polynomial regression, which Holt's double exponential smoothing reached MAPE 0.056% in the training set, 0.047% in one-step testing, and 0.898% in multi-step testing. The 2-dimensional chart also reported that Holt's double exponential smoothing was more suitable than polynomial regression to make intersection points between actual data and forecast result. We concluded that the forecasting model generated by Holt's Double Exponential Smoothing is feasible to be applied to daily forecasting for gold bullion prices.

REFERENCES

- Aslam, S., Heijden, G. V., & Collins, H. (2019). Predicting customer loyalty using various regression models. *Advanced Machine Learning*, 21, 1-11.
- 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.
- Basov, V. (2015). The world's highest grade gold mines. Retrieved from <https://www.mining.com/the-worlds-highest-grade-gold-mines/>
- Bose R., Dey R.K., Roy S., & Sarddar D. (2020). Time series forecasting using double exponential smoothing for predicting the major ambient air pollutants. In M. Tuba, S. Akashe, & A. Joshi (Eds.), *Information and communication technology for sustainable development*. Singapore: Springer.
- Djatmiati, T. S., Alfath, T. P., & Fahrudin, T. M. (2019). Forecasting public services maladministration in local government. *Proceedings of the 1st International Conference on Business, Law and Pedagogy (ICBLP) 2019*, 1-11.
- Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing*. Berlin: Springer-Verlag Berlin Heidelberg.
- Karanam, M., & Shenbagavalli, R. (2019). Investment pattern of millennials: A path ahead. *Research Journal of Social Science and Management*, 9(4), 24-29.
- Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669-679.
- Mombeini, H., & Yazdani-Chamzini, A. (2015). Modeling gold price via artificial neural network. *Journal of Economics, Business and Management*, 3(7), 699-703.
- Mulyana, F., & Djajadiningrat, S. T. (2013). Maximizing the contribution of mining industry to sustainable development: Case study at PT Freeport Indonesia. *The Indonesian Journal of Business Administration*, 2(19), 2349-2358.

- Nazim, A., & Afthanorhan, A. (2014). A comparison between single exponential smoothing (SES), double exponential smoothing (DES), holt's (brown) and adaptive response rate exponential smoothing (ARRES) techniques in forecasting Malaysia Population. *Global Journal of Mathematical Analysis*, 2(4), 276-280.
- Peddada, S. D., & Haseman, J. K. (2005). Analysis of nonlinear regression models: A cautionary note. *Dose-Response: A Publication of International Hormesis Society*, 3(3), 342-352.
- Pusca, A. C., Tureac, C. E., Andronic, B., & Filip, A. C. (2012). Gold evolution and analysis on the commodity exchange. *International Journal of Management Cases*, 14(1), 71-78.
- PwC. (2018). Mining in Indonesia: Investment and taxation guide (10th ed.). Indonesia: PwC Indonesia.
- Singawinata, I. P. (2005). The future of the Indonesian mining industry: Recommendations to policy makers. *Ritsumeikan Journal of Asia Pacific Studies*, 22, 99-113.
- Street, L. (2019). What do millennial investors think of gold? Retrieved from <https://www.gold.org/goldhub/gold-focus/2019/06/what-do-millennial-investors-think-gold>
- Tyrrell, T. J. (1983). The use of polynomials to shift coefficients in linear regression models. *Journal of Business and Economic Statistics*, 1(3), 249-252.
- Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. *Procedia - Social and Behavioral Sciences*, 106, 234-240.
- Wang, H., Sheng, H., & Zhang, H. (2019). Influence factors of international gold futures price volatility. *Transactions of Nonferrous Metals Society of China*, 29(11), 2447-2454.