

Final Report: Team 8

Predicting Prices of Oil and Gold

Ayush Sengupta, Benjamin Lin, Komal Sanjeev, and Sreevathsan
Ravichandran

Department of Computer Science, Stony Brook University,
Stony Brook, NY 11794-4400
{aysengupta,xianlin,ksanjeev,sravichandra}@cs.stonybrook.edu
<http://www.cs.stonybrook.edu/~skiena/591/projects>

1 Challenge

Our challenge is to predict the prices of Oil and Gold on January 1st 2015 as of December 1st in 2014 (a month in advance).

1.1 Oil

Oil is a non-renewable resource which occurs in the earth. It is extracted and sent to refineries where different petroleum products such as gasoline, petrol, heating oil, etc are separated. Almost two-thirds of our energy demands are met by oil. Oil is the most heavily and actively traded commodity, which accounts to almost 10% of the world's trade.

The price of oil is determined by supply and demand. An increase in demand results in an increase in the price of oil. The supply and demand themselves are determined by various factors such as economy, weather and geopolitics.

Being an extensively traded commodity, coupled with several factors which affect its price, the oil price is very volatile. Some of the factors might even be interdependent, which makes it very difficult to estimate the extent to which an individual factor could affect the price. Moreover, the relationship between certain factors and the oil prices could also change over time. These factors make the prediction of oil prices a very complex and challenging problem.

1.2 Gold

Gold is viewed as a symbol of wealth since the ancient times of the human history. It is widely used as jewelry and in some precision gages.

For the long period of time, the gold price is fixed; until 1968, with the breakdown of Bretton Woods Currency Arrangement, the gold price started to be market determined. People also invest in gold, especially when the melt-down

of US dollars, because gold plays a significant role as a stabilizing influence for investment portfolio.

The gold price faces similar challenges as the oil price. It is dynamic and highly volatile. Multiple factors affect the gold price. The extent of correlation and the level of importance of each factor are not clearly known. Also similar to the difficulties we face in oil price prediction, some of these factors are interdependent, so we need to estimate the level of inter-correlation between them.

2 History/Background

2.1 Factors Affecting the Price of Oil

Crude oil prices are determined by the balance between supply and demand. An increase (or decrease) in demand causes the price of oil to rise (or fall). Consequently, a cutback in the production of oil results in an increase in oil prices. There are several factors which disturb the supply-demand balance, thus resulting in oil price fluctuations.

Organization of the Petroleum Exporting Countries (OPEC) Supply

OPEC is an organization of 12 oil exporting nations, namely Algeria, Angola, Ecuador, Iran, Iraq, Kuwait, Libya, Nigeria, Qatar, Saudi Arabia, United Arab Emirates, and Venezuela. It aims at coordinating and unifying petroleum prices of its member countries [1]. Oil supply from the OPEC member countries represents about 40% of the world's crude oil, and their actions can affect the prices of oil to a significant extent. For example, limiting the oil production from OPEC's major oil producers such as Saudi Arabia can influence crude oil supply and affect the prices [2].

Oil prices not only depend on the current demand and supply, but also on the projected future supply and demand. OPEC adjusts the oil productions of its member countries based on current and future demand.

Non-OPEC Supply The non-OPEC countries produce about 60% of the world's crude oil. A lack of supply from the non-OPEC countries creates additional pressure on the OPEC countries which can also contribute to a rise in oil prices [2].

Stock Price As economic conditions improve, there is an increase in demand for several commodities including oil, which results in an increase in oil prices. The Standard and Poor's 500 (S&P 500) index is a common benchmark for the stock market of USA. It is a weighted index of the market capitalization of 500 companies. It is the most commonly used indicator of the US economy.

Market Transactions - Spot price and Futures The current price at which a commodity can be traded at a specific place and time is called the spot price. Crude oil can be purchased on the spot at the current market price.

Crude oil can also be traded in the futures market. A futures market is when a commodity is traded in futures contracts, i.e., a contract between a buyer and seller where the buyer agrees to buy a certain quantity of a commodity for a fixed price at a time in the future.

The uncertainty of futures contract prices tend to affect the spot oil prices.

Seasonal Effects Certain crude oil products such as heating oil and gasoline tend to have a seasonal variance. For example, there is an increase in demand for oil in the fourth quarter due to the cold weather, and a subsequent reduction in demand during late winter as the weather gets warmer. Gasoline prices also tend to rise in the summer due to an increased consumption.

US Dollar Exchange Rate (USD_X) Oil is traded in Dollars, and thus, any change in the Dollar exchange rate relative to other currencies can cause oil prices to shift. Several studies support the negative correlation between the dollar exchange rate and the price of oil. There are several arguments to support this statement. One reason could be that the depreciation in the dollar exchange rate causes the oil to be cheaper in countries outside the US, thus leading to an increase in demand, which in turn causes oil prices to rise. However, it has been observed that the relationship between oil prices and the dollar exchange rate has not been stable over the years [2] [3].

2.2 Factors Affecting the Price of Gold

Gold is a precious metal which was used as a currency in several major civilizations in the past centuries. In the United States of America, gold had been at a fixed price of about \$20 per ounce, since the early 19th century. In 1934, President Franklin Roosevelt raised gold price from \$20.67 per ounce to \$35. In 1968, with the breakdown of Bretton Woods Currency Arrangements, the gold price became market determined. From Figure 1 below, we observe that the two peaks of the gold price coincide with two significant economic recessions in our history - one in 1980, and the other 2008.

Based on our paper review, we found the following 9 factors would affect the gold price: Crude Oil Price, Commodity Research Bureau future index (CRB), EURO/USD Foreign Exchange Rate (EURO/USD), Inflation rate (INF), Money Supply (M1), and US Dollar index (USD_X) [4][5][6]. In the model development process, we selected some of them to be included in our models.



Fig. 1. Gold price in dollars per ounce from Jan 1, 1970 to Dec 9, 2014

Inflation In history, during inflation, people tended to keep precious commodities such as gold instead of paper money to ensure their assets were not reduced too much. Therefore, the higher the inflation is, the higher the gold price is in general.

Stock Price Similar to the oil price case, NYSE and S&P 500 indexes are common benchmarks for the US economy.

US Dollar Exchange Rate (USD_X) Similar to the oil price case, US dollar is the currency in trading gold. The raise of the USD_X implies an increase of the US dollar purchasing power, which reduces the consumers' motivation of keeping gold, and possibly decreasing the gold price.

Consumer Sentiment Index (CSI) The Consumer Sentiment Index or Consumer Confidence Index, is an index measuring the consumers' confidence over the market [9][10]. Similar to the effects of USD_X over the gold price, a high CSI indicates consumers generally feel optimistic about the overall economy and their ability of obtaining and keeping their jobs, so they would be less likely to keep precious metals like gold.

Crude Oil Price Crude oil prices have an approximate 85% to 92% correlation with gold prices. Rising oil prices may lead to an increase in the gold price, but the converse may not be true [4][5]. The gold-oil relation suggests that the crude oil price could partly account for inflation. An increase in the oil price results in increased prices of gasoline. Gasoline being more expensive results in an increase in the cost to transport goods, thus causing a possible hike in prices of goods. The final result is an increased price level in other words, inflation. Gold tends to appreciate with inflation. Therefore, elevated oil prices can eventually lead to higher gold prices [7].

3 Literature Review

There is a significant amount of papers aiming on understanding and predicting oil and gold prices. Various studies have developed different prediction models based on different techniques and factors. Some studies try to make predictions based on historical oil and gold prices. Others focus on the economic aspects and try to explain the correlations between the prices of oil and gold with respect to a variety of economic factors. Therefore, we evaluate and summarize some widely used models into the following two categories: Standard Time Series Models and Structural Models Considering Economic Factors.

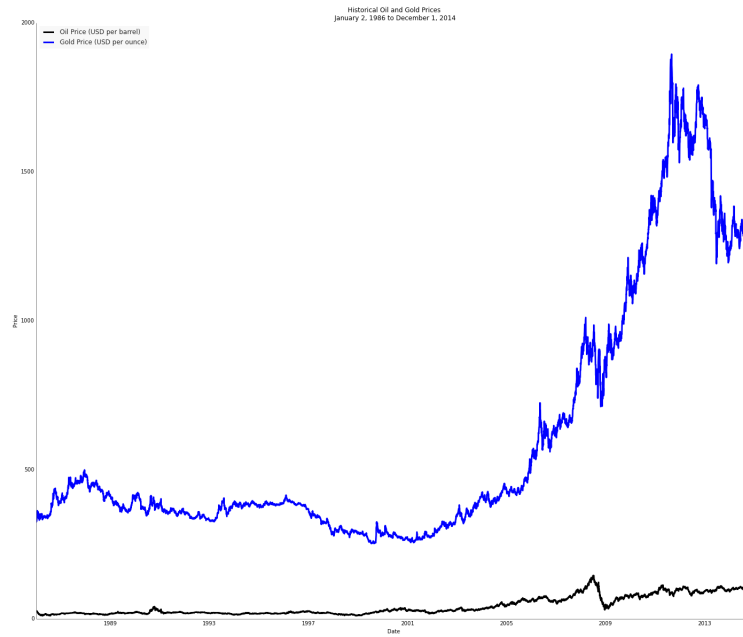


Fig. 2. Historical Oil and Gold Prices (1986-2012) show a correlation, particularly before 2009

3.1 Standard Time Series Models

Standard time series models attempt to predict the oil price using the current and historical oil prices. The same strategy applies to predicting the gold price as well. This type of models is useful in the following conditions:

1. The prices show autocorrelation and autoregressive behavior, i.e., there is a pattern or a significant correlation between current and the previous prices.
2. There are a large number of explanatory variables and it is difficult to understand them well because they interact with each other in a very complicated manner.
3. Forecasting the dependent variable may require predicting the explanatory variables. And prediction of the explanatory variable might in turn be a harder problem.
4. Not all explanatory factors and variables are known.

The most basic time series models that have been applied to model oil prices are the autoregressive models. In general, an autoregressive model $\mathbf{AR}(\mathbf{p})$ tries to model the current value of a time series based on the value of the last 'p' instances in the time series.

$$\text{Thus, } X_t = c + \sum_{i=1}^p \varphi_{t-i} X_{t-i} + \epsilon_t.$$

The co-efficients are regressed to predict the current price.

Similarly, Autoregressivemoving-average (ARMA) models take in to account the moving average factor. They try to predict the randomness based on the historical prices. Hence $\mathbf{ARMA}(\mathbf{p}, \mathbf{q})$ can be written as:

$$X_t = c + \sum_{i=1}^p \varphi_{t-i} X_{t-i} + \epsilon_t + \sum_{i=1}^q \theta_{t-i} X_{t-i}.$$

Further, it is known that oil price changes(volatility) follow GARCH/ARCH properties. Least square models generally assume that the expected value of all error terms, when squared, is constant. This assumption is termed as homoskedasticity. Data(time series) in which this conditions fail to hold, are heteroskedastic. ARCH and GARCH models treat heteroskedasticity as a variance which is then modelled autoregressively.[11]

In short, GARCH models split the error-terms ε_t into a stochastic component z_t and a time dependent variance σ_t^2 . Thus, $\varepsilon_t = z_t + \sigma_t$. The series σ_t^2 in $\mathbf{ARCH}(\mathbf{q})$ is modelled by:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2$$

Similarly **GARCH**(**p,q**) is modelled by:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2$$

GARCH and ARCH models have consistently been used in the literature to predict oil prices with varying degrees of accuracy.

Particularly for predicting the gold price, Hadavandi et al.(2010) provides a time series model using Particle Swarm Optimization(PSO) for parameter estimation [12]. The method is illustrated as following:

$$GP(k) = \alpha_1 GP(k-1) + \alpha_2 GP(k-2) + \alpha_3$$

where GP(k) is Gold price in kth day is given by equation

PSO is an evolutionary computation technique based on the simulation of bird flocking in 2 dimensions, which is used to estimate the parameters α_1 , α_2 , α_3 .

$$GP(k) = 0.976 * GP(k-1) + 0.1373 * GP(k-2) + 0.1157$$

They show that their method performed better than ARIMA, ANN and OLS methods[13]. The following table compares the Mean Absolute Error (MAE) of the proposed model with models proposed by Khashei et al. [14] and OLS method.

Method	MAE
Hybrid of ANN and Fuzzy Regression [14]	0.097
ARIMA [14]	0.105
Hybrid of ARIMA, ANN and Fuzzy [13]	0.065
OLS method	0.098
PSO-based time series model(Proposed model)	0.047

Hadavandi's method, however, cannot guarantee to give good results for long term price predictions because they only use the gold price of the last three days. Nevertheless, since it was a simple model which gave accurate results for the short-term, we implemented it as one of our baseline models for our first report.

3.2 Structural Models Considering Economic Factors

For the price of oil, structural models consider the oil price to be modelled as a function of certain explanatory variables such as oil consumption and production, OPEC behaviour, interest rates, exchange rates, and other commodity prices. The major drawback of using structural models to predict oil prices is that the models are extremely complex, and there is a strong inter correlation between factors themselves. Hence, there have not been many studies that focus on structural analysis to forecast oil prices.

According to Huntington(1994)[15], structural demand and supply models are generally not successful in predicting oil prices due to inaccurate forecasts of GDP and the oil supply from different countries. Another reason was not taking into account the market participation expectation of OPEC countries.

However, some interesting work has been done based solely on structural models. Most of these studies use models and results that instead of trying to predict the price, try to understand the nature of the oil market. Also, these models predict short-term oil prices, and it is unclear if they could be used for long term forecasting. One such interesting study by Pindyck(1999)[16] shows that long term oil prices are mean reverting around shifting trend lines.

In another direction Yang et al.(2002) [17] introduces a model to determine the factors affecting US oil prices. First, they highlight the unstable demand structure of the oil market. Then, they use a GARCH model (general autoregressive conditional heteroskedasticity) to investigate the volatility of oil prices. Using the co-efficients they generated, they estimate that the future oil price will be 0.987 times the current oil price if the US GDP decreases by 5%.

Similarly, structural models considering a variety of economic factors also apply to the gold price prediction. According to Ismail et al.(2009)[6], they design models with the gold price as the only dependent variable, alongside different numbers of independent variables. Initially, they propose that the gold price is dependent on the following 8 factors: Commodity Research Bureau future index (CRB), USD/Euro Foreign Exchange Rate (EUROUSD); Inflation rate (INF); Money Supply (M1); New York Stock Exchange (NYSE); Standard and Poor 500 (SPX); Treasury Bill (T-BILL) and US Dollar index (USDIX). Therefore, their first-order regression model, which they call it naive model, is like:

$$\hat{Y} = -560.618 + 0.712X_1 + 161.740X_2 - 7.836X_3 + 0.424X_4 - 0.010X_5 + 0.010X_6 + 3.198X_7 + 0.580X_8$$

where \hat{Y} is the predicted gold price; X_1 is CRB; X_2 is EUROUSD; X_3 is INF; X_4 is M1; X_5 is NYSE; X_6 is SPX; X_7 is T-Bill; X_8 is USDIX.

Then they show that using stepwise regression, the number of independent variables can be reduced from 8 to 4. Their enhanced model is

$$\hat{Y} = -301.509 + 0.676X_1 + 114.651X_2 - 5.563X_3 + 0.309X_4$$

where \hat{Y} is the predicted gold price, X_1 is CRB; X_2 is EUROUSD; X_3 is INF; X_4 is M1.

Ismail's paper provides us the intuition of what factors we may use for building our own advanced models (autoregressive and multiple linear regression models) in predicting the gold price. In our advanced models, we include the S&P 500 Index, NYSE Index and the US Dollar Index that are mentioned in Ismail's paper, as well as Consumer Sentiment Index and the oil price that are not. The details will be discussed in Section 7 "Advanced Models".

4 Data Sets

Our data consists of multiple time series of monthly Oil and Gold prices[18], and the following macroeconomic factors:

- S&P 500 Index [18]
- New York Stock Exchange Index (NYSE) [18]
- US Dollar Index [18]
- Consumer Sentiment Index (CSI) [19]

Date	Oil Price	Gold Price	S&P 500	NYSE	USD Index	CSI
11/30/2014	75.74	1182.8	2067.56	10955.41	80.81	86
10/31/2014	80.53	1164.3	2018.05	10845	80.8143	86
9/30/2014	91.17	1216.5	1972.29	10702.93	81.0908	86
8/31/2014	97.86	1285.8	2003.37	11046.29	77.9769	93.4
7/31/2014	98.23	1285.3	1930.67	10726.43	77.2128	90.3
6/30/2014	106.07	1315	1960.23	10979.42	75.7271	86.4

Fig. 3. Data frame for oil and gold price and its related macroeconomic factors

Before calculating the heat map of correlation for oil and gold prices, we have all the prices and macroeconomic factors inflation adjusted.

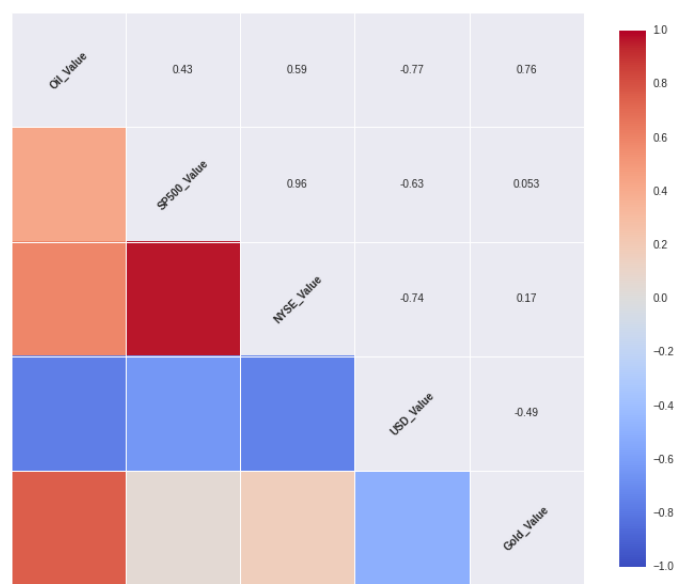


Fig. 4. Correlation Heat Map for Oil Price and related macroeconomic factors

	<i>S&P 500</i>	<i>NYSE</i>	<i>USD Index</i>	<i>Gold Price</i>
<i>OilPrice</i>	0.43	0.59	−0.77	0.76

Table 1. Correlation Matrix - Oil

	<i>S&P500</i>	<i>NYSE</i>	<i>USD</i>	<i>CSI</i>	<i>OilPrice</i>
<i>GoldPrice</i>	0.053	0.17	−0.49	−0.54	0.76

Table 2. Correlation Matrix - Gold

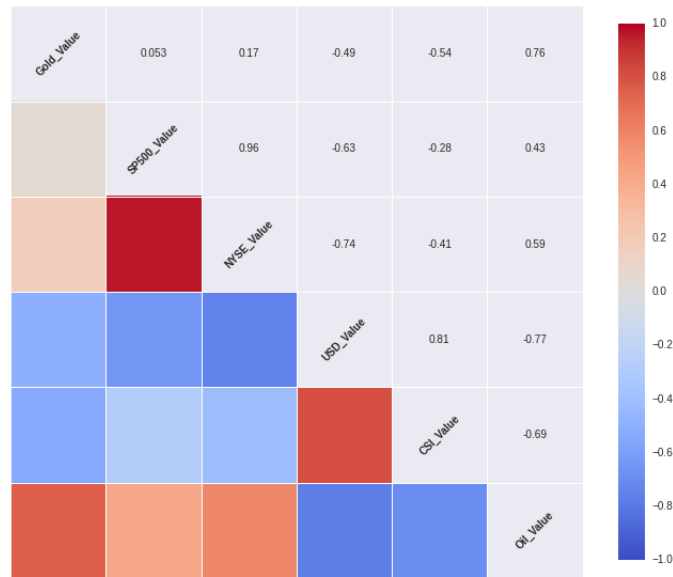


Fig. 5. Correlation Heat Map for Oil Price and related macroeconomic factors

The correlation heat map in Figure 4 and Figure 5, and the corresponding tables containing correlation coefficients show the correlation between the price of oil/gold and related economic factors.

This time, we have all the values of the two commodity prices and necessary macroeconomic factors adjusted by inflation. For the oil price, the correlation between the oil price and S&P 500 is dropped from 0.72 to 0.43, the correlation between the oil price and NYSE index is dropped from 0.81 to 0.59, and the correlation between the oil price and gold price is dropped from 0.87 to 0.76. Similarly, for the gold price, the correlation between the gold price and S&P 500 is dropped from 0.53 to 0.053, the correlation between the gold price and NYSE index is dropped from 0.59 to 0.17, the absolute value of the correlation between the gold price and the US dollar index is dropped from 0.75 to 0.49, and the correlation between the gold price and oil price is dropped from 0.87 to 0.76. By doing inflation adjustment, we are certain to have increased the predicting power of our autoregressive with multiple linear regression models. For details, refer to Section 7 Advanced Models.

Our goal is to predict the prices of oil and gold on January 1st, 2015 as of December 1st, 2014, which is a month in advance. Therefore, we have used monthly data, i.e., oil/gold price and values of other economic factors on the last day of every month. But this severely limits the amount of data we can obtain. Although we collected data for the past 30 years, we have just 360 entries in our time series, which isn't a good enough number to generate good enough error metrics. This also limits our prediction power.

5 Observations

Report on interesting things you can learn/visualize from your data set. Include and describe these visualizations. (2-4 pages)

6 Baseline Models

In order to illustrate an improvement in the accuracy of our predictions, we compare our advanced models (see Section 7 Advanced Models for details) to the following two baseline models:

- **Baseline Model 1** The price is the same as the previous month's price:

$$P_t = P_{t-1}$$

- **Baseline Model 2** The price is a cubic weighted average of the last 3 months:

$$P_t = \frac{1}{\{k(k+1)\}^2} \sum_{i=1}^k (k-i+1)^3 P_{t-i}$$

Error Metrics of Baseline Models for Oil The following are the error metrics and the error histograms for our baseline models for predicting the oil price:

Model	Relative Error	Mean Absolute Error	RMSE	Std Dev of Err %	Probability
<i>Baseline Model 1</i>	6.58	5.93	7.85	0.0904	0.7851
<i>Baseline Model 2</i>	7.17	6.37	8.58	0.1025	0.8037

Table 3. Error Metrics for Baseline Model - Oil. RMSE stands for Root Mean Squared Error; Probability refers to the probability that the error is within 1 standard deviation

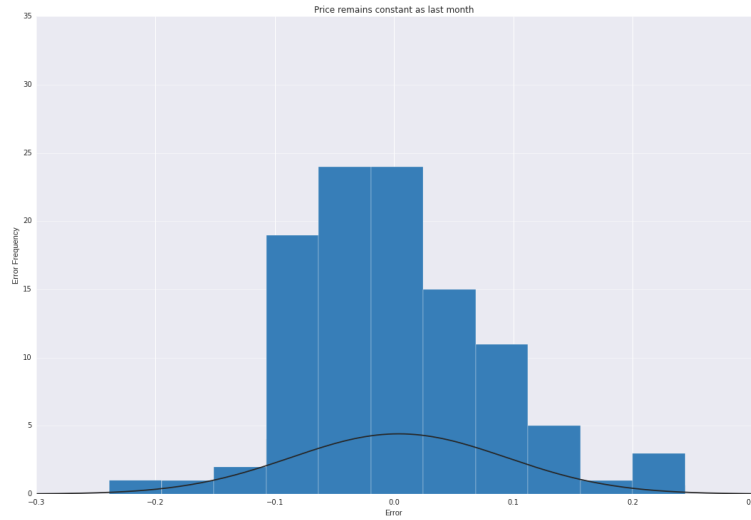


Fig. 6. Error plot for Baseline Model 1 - Oil price is the same as the previous month's price

The predicted price of WTI Crude Oil on Jan 1, 2015 as of Dec 1, 2014 is **75.74 USD per Barrel** based on Baseline Model 1.

The predicted price of WTI Crude Oil on Jan 1, 2015 as of Dec 1, 2014 is **76.99 USD per Barrel** based on Baseline Model 2.

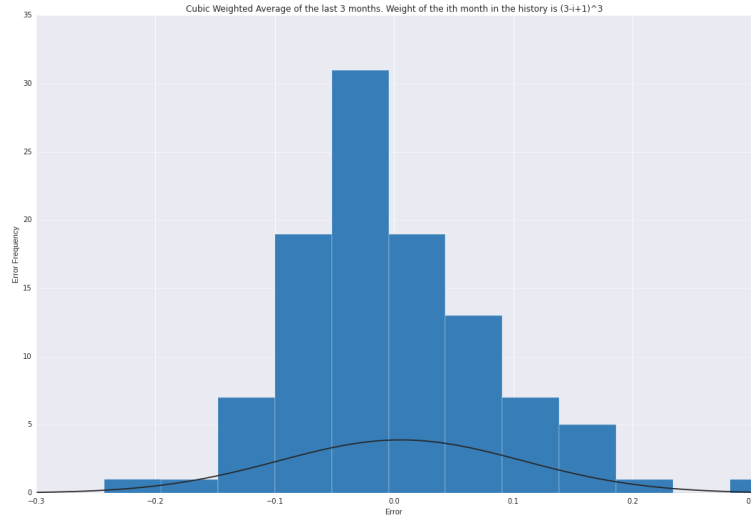


Fig. 7. Error plot for Baseline Model 2 - Oil price is a cubic weighted average of the last 3 months

Error Metrics of Baseline Models for Gold The following are the error metrics and the error histograms for our baseline models for predicting the gold price:

Model	Relative Error	Mean Absolute Error	RMSE	Std Dev of Err %	Probability
<i>Baseline Model 1</i>	4.61	56.11	75.36	0.0603	0.7196
<i>Baseline Model 2</i>	4.56	55.78	73.27	0.0586	0.6916

Table 4. Error Metrics for Baseline Model - Gold. RMSE stands for Root Mean Squared Error; Probability refers to the probability that the error is within 1 standard deviation

The predicted price of Gold Jan 1, 2015 as of Dec 1, 2014 is **1182.80 USD per ounce** based on Baseline Model 1.

The predicted price of Gold Jan 1, 2015 as of Dec 1, 2014 is **1179.84 USD per ounce** based on Baseline Model 2.

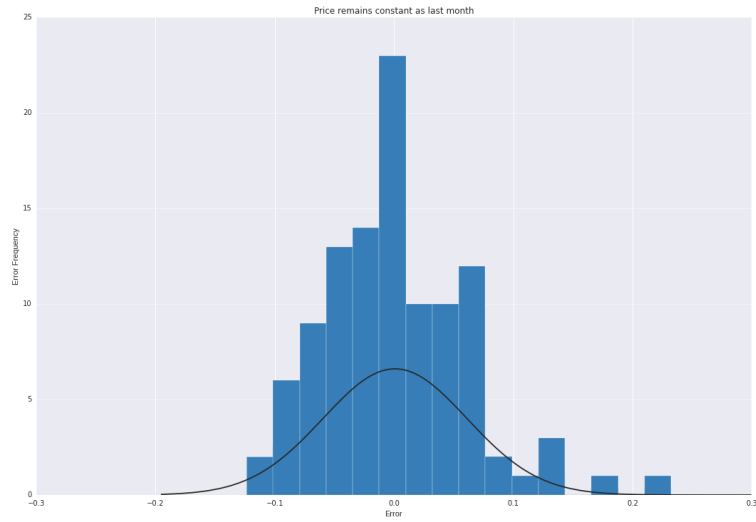


Fig. 8. Error plot for Baseline Model 1 - Gold price is the same as the previous month's price

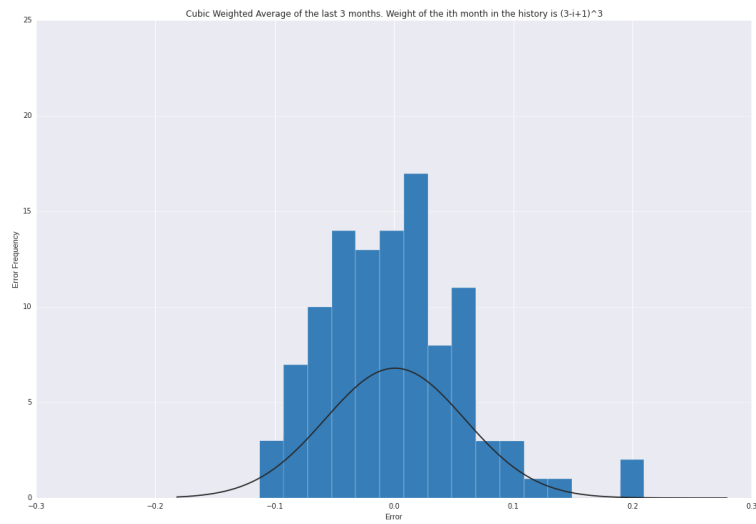


Fig. 9. Error plot for Baseline Model 2 - Gold price is a cubic weighted average of the last 3 months

7 Advanced Models

Describe your development of one or more advanced models (presumably machine learning-based), and present results from your evaluation environment showing the performance of it against your baseline models. What methods of fitting/data analysis are you using? (3-5 pages)

8 Final Prediction and Conclusions

State your final forecast for your challenge. Discuss any difficulties you had to overcome in building a good model, and fruitful investigations for subsequent groups. (1-3 pages)

Acknowledgments. Here acknowledge any other people who helped with this project.

9 Bibliography

References

1. "OPEC : Home.". <http://www.opec.org/>
2. "Energy and Financial Markets: What Drives Crude Oil Prices". <http://www.eia.gov/finance/markets>
3. Zhang, Yue-Jun, et al. "Spillover effect of US dollar exchange rate on oil prices." *Journal of Policy Modeling* 30.6 (2008): 973-991.
4. Shahriar Shafiee, Erkan Topal.:An overview of global gold market and gold price forecasting. *ScienceDirect*. Volume 35, Issue 3, September 2010, 178–189 (2010)
5. Zhang, Yue-Jun, and Yi-Ming Wei. "The crude oil market and the gold market: Evidence for cointegration, causality and price discovery." *Resources Policy* 35.3 (2010): 168-177.
6. Z. Ismail, A. Yahya and A. Shabri. "Forecasting Gold Prices Using Multiple Linear Regression Method." *American Journal of Applied Sciences* 6 (8) (2009): 1509-1514.
7. The Relationship between Gold and Crude Oil Price, <http://www.marketoracle.co.uk/Article38141.html>
8. What's the difference between consumer confidence and consumer sentiment?, <http://www.investopedia.com/ask/answers/09/consumer-confidence-sentiment-difference.asp>
9. What's the difference between consumer confidence and consumer sentiment?, <http://www.investopedia.com/ask/answers/09/consumer-confidence-sentiment-difference.asp>
10. Consumer confidence index, http://en.wikipedia.org/wiki/Consumer_confidence_index
11. Engle, Robert. "GARCH 101: The use of ARCH/GARCH models in applied econometrics." *Journal of economic perspectives* (2001): 157-168.

12. Hadavandi, Esmail, Arash Ghanbari, and Salman Abbasian-Naghneh. "Developing a Time Series Model Based on Particle Swarm Optimization for Gold Price Forecasting." Business Intelligence and Financial Engineering (BIFE), 2010 Third International Conference on. IEEE, 2010.
13. Khashei, Mehdi, Mehdi Bijari, and Gholam Ali Raissi Ardali. "Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs)." Neurocomputing 72.4 (2009): 956-967.
14. Khashei, Mehdi, Seyed Reza Hejazi, and Mehdi Bijari. "A new hybrid artificial neural networks and fuzzy regression model for time series forecasting." Fuzzy Sets and Systems 159.7 (2008): 769-786.
15. Huntington, Hillard G. "Oil price forecasting in the 1980s: what went wrong?." The Energy Journal (1994): 1-22.
16. Pindyck, Robert S. "The long-run evolution of energy prices." The Energy Journal (1999): 1-27.
17. Yang, C. W., Ming-Jeng Hwang, and Bwo-Nung Huang. "An analysis of factors affecting price volatility of the US oil market." Energy Economics 24.2 (2002): 107-119.
18. Quandl - Find, Use and Share Numerical Data. <https://www.quandl.com>
19. http://future.aae.wisc.edu/data/monthly_values/by_area/998?grid=true