Final Project - Forecasting Gold Price

Aliaksandr Panko

February 21, 2018

Contents

1 Objectives

To forecast gold price evolution we would like to compare a linear regression estimated by the Ordinary Least Squares (OLS) method and another model.

2 Data

Gold prices are given:

- Data Source

- Frequency: Annual data
- Period: 1978 - 2014

3 Linar Regression

In the first part of the project OLS linear regression is used.

1. Explanatory variable: Previous gold price
2. Explained variable: Current gold price

The regression equation is:

$$gold_t = \alpha + \beta gold_{t-1}$$

3.1 Results of the regression

Original:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   gold   R-squared:                       0.906
Model:                            OLS   Adj. R-squared:                  0.903
Method:                 Least Squares   F-statistic:                     326.3
Date:                Tue, 20 Feb 2018   Prob (F-statistic):           5.31e-19
Time:                        22:31:23   Log-Likelihood:                -222.73
No. Observations:                  36   AIC:                             449.5
Df Residuals:                      34   BIC:                             452.6
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         36.1122     35.585      1.015      0.317     -36.205     108.429
gold_lag       0.9882      0.055     18.064      0.000       0.877       1.099
==============================================================================
Omnibus:                        5.980   Durbin-Watson:                   1.175
Prob(Omnibus):                  0.050   Jarque-Bera (JB):                4.617
Skew:                           0.654   Prob(JB):                       0.0994
Kurtosis:                       4.169   Cond. No.                     1.15e+03
==============================================================================
```

Logarithm:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               log_gold   R-squared:                       0.869
Model:                            OLS   Adj. R-squared:                  0.865
Method:                 Least Squares   F-statistic:                     225.0
Date:                Tue, 20 Feb 2018   Prob (F-statistic):           1.48e-16
Time:                        22:32:31   Log-Likelihood:                 8.9141
No. Observations:                  36   AIC:                            -13.83
Df Residuals:                      34   BIC:                            -10.66
Df Model:                           1
Covariance Type:            nonrobust
================================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const          0.4074      0.386      1.056      0.299      -0.377       1.192
log_gold_lag   0.9420      0.063     15.001      0.000       0.814       1.070
==============================================================================
Omnibus:                       10.138   Durbin-Watson:                   1.234
Prob(Omnibus):                  0.006   Jarque-Bera (JB):                9.400
Skew:                           0.993   Prob(JB):                      0.00910
Kurtosis:                       4.523   Cond. No.                         75.1
==============================================================================
```

As we can see in both models constant is not significant at 95% significance level (p_value > 0.05)

Based on R^2 we can conclude that in both cases the model describes gold prices evaluation rather good(gold price in the previous period explains 90.56 % and 86.87 % of today's evolution respectively)

So, it is a good idea to implement this model without intercept.

3.2 Model Without Alpha

The regression equation is:

$$gold_t = \beta gold_{t-1}$$

This model gives us next result:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   gold   R-squared:                       0.969
Model:                            OLS   Adj. R-squared:                  0.969
Method:                 Least Squares   F-statistic:                     1110.
Date:                Tue, 20 Feb 2018   Prob (F-statistic):           4.23e-28
Time:                        22:45:16   Log-Likelihood:                -223.26
No. Observations:                  36   AIC:                             448.5
Df Residuals:                      35   BIC:                             450.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
gold_lag       1.0340      0.031     33.313      0.000       0.971       1.097
==============================================================================
Omnibus:                        3.876   Durbin-Watson:                   1.193
Prob(Omnibus):                  0.144   Jarque-Bera (JB):                3.211
Skew:                           0.182   Prob(JB):                        0.201
Kurtosis:                       4.417   Cond. No.                         1.00
==============================================================================
```

The result is extremely good, 97% of gold price evaluation can be explained by this model! However Durbin-Watson test and Jarque-Bera test signal that there is autocorrelation problem and residuals are not normally distributed.

Actually, it is well-known problem: "since stock prices tend not to change too radically from one day to another, the prices from one day to the next could potentially be highly correlated, even though there is little useful information in this observation. In order to avoid autocorrelation issues, the easiest solution in finance is to simply convert a series of historical prices into a series of percentage-price changes from day to day. " ref. Investopedia

In a conclusion, I would rather use model with percentage that are mentioned above.

3.3 Multiple regression

I decided to take into consideration next factors:

1. Producer Price Index for All Commodities [PPIACO]
2. Wilshire 5000 Total Market Index [WILL5000IND]

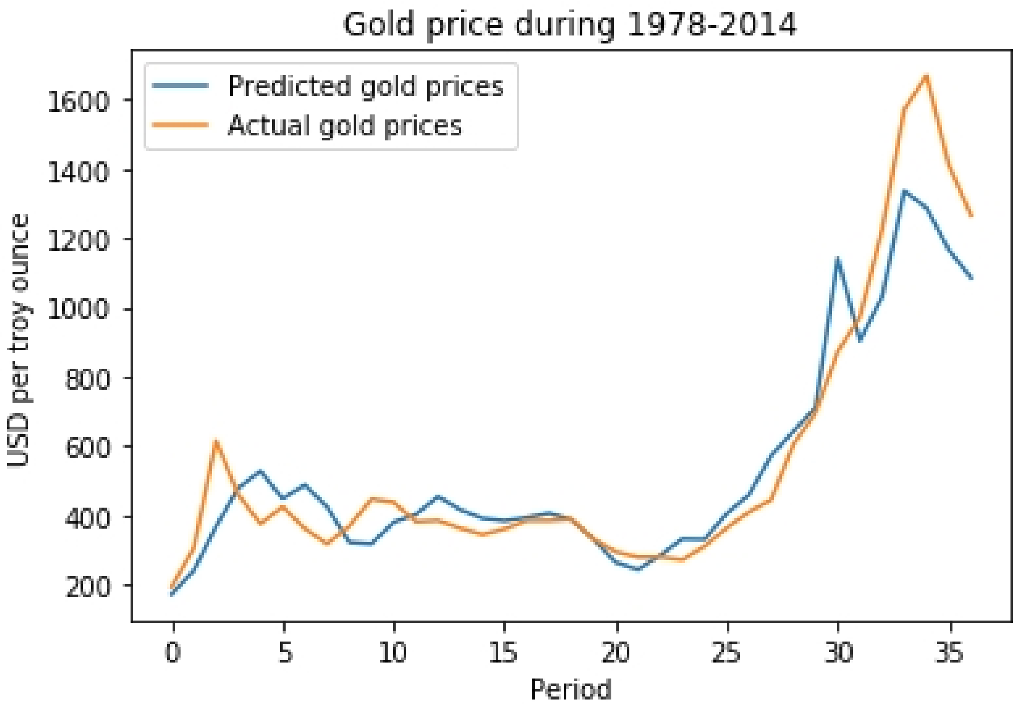
Frankly speaking, it was not so easy to find out these factors. I also tried money supply and interest rates but they were not significant.

Likely in the end the model gives rather good result:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.878
Model:                            OLS   Adj. R-squared:                  0.871
Method:                 Least Squares   F-statistic:                     122.6
Date:                Wed, 21 Feb 2018   Prob (F-statistic):           2.83e-16
Time:                        00:10:59   Log-Likelihood:                 9.5962
No. Observations:                  37   AIC:                            -13.19
Df Residuals:                      34   BIC:                            -8.360
Df Model:                           2
Covariance Type:            nonrobust
===============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
const         -11.5109      1.293     -8.905      0.000     -14.138      -8.884
PPIACO          3.9755      0.301     13.208      0.000       3.364       4.587
WILL5000IND    -0.6026      0.069     -8.779      0.000      -0.742      -0.463
==============================================================================
Omnibus:                        1.471   Durbin-Watson:                   1.014
Prob(Omnibus):                  0.479   Jarque-Bera (JB):                0.910
Skew:                           0.382   Prob(JB):                        0.634
Kurtosis:                       3.073   Cond. No.                         237.
==============================================================================
```

As we can see all factors are significant at 1% level and 87.8% data are explained by the model. Nevertheless Durbin-Warson test still is not perfect that means autocorrelation existence.

Let us compare the results visually:



As we can see it is good enough.

4 ARMA model

Since we have rather good prediction taking into account only previous value I decided to try ARMA model model for predicting gold price evolution.

4.1 Stationarity

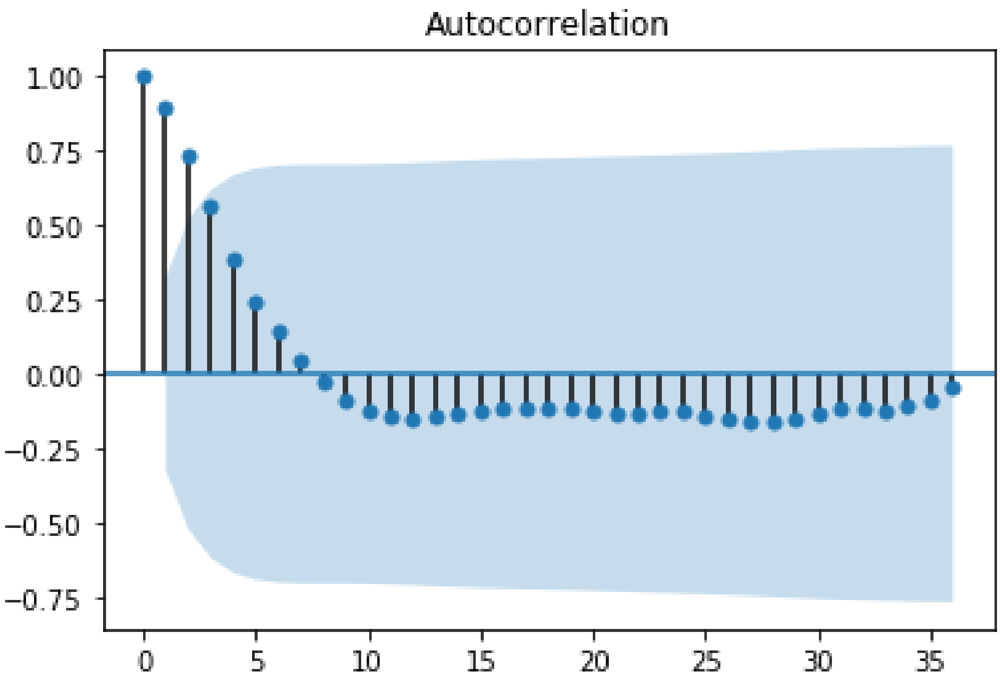
Firstly, we need to check stationarity. For this purposes I used Augmented Dickey-Fuller unit root test.

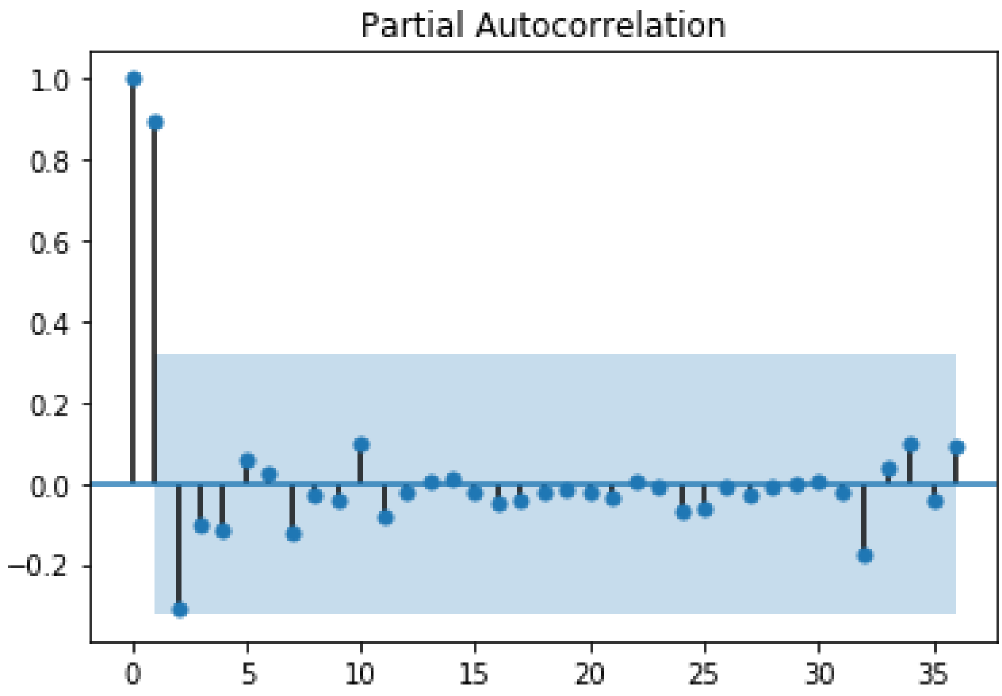
The null hypothesis of the **Augmented Dickey-Fuller** is that there is a unit root, with the alternative that there is no unit root.

p_value is 3.13184167501e-06 - this means that we reject H0 and assume that the **process is stationary**.

4.2 ACF & PACF

Secondly, we need to check ACF and PACF and use Box–Jenkins method.



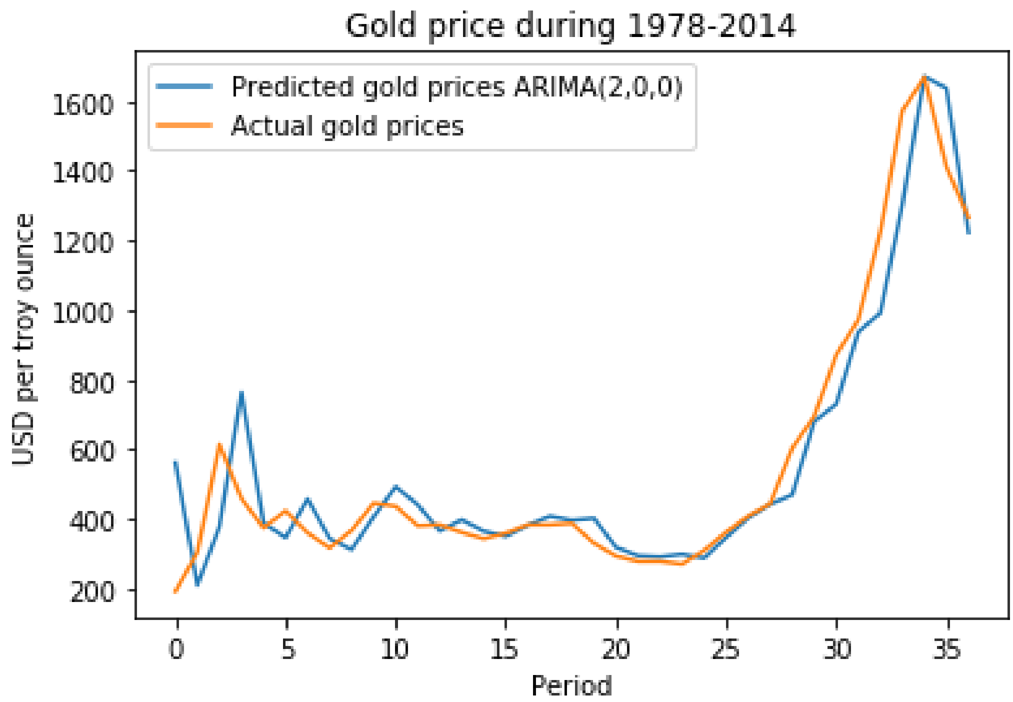


It's clear that ACF decays exponentially and PACF has 2 significant spikes. Since differencing is not required, the parameters of the model are **ARIMA(2,0,0)**.

Let us have a look at the results:

```
                              ARMA Model Results
==============================================================================
Dep. Variable:                      y   No. Observations:                   37
Model:                     ARMA(2, 0)   Log Likelihood                -226.846
Method:                       css-mle   S.D. of innovations            106.968
Date:                Wed, 21 Feb 2018   AIC                            461.692
Time:                        01:02:23   BIC                            468.136
Sample:                             0   HQIC                           463.964

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const        562.3753    214.407      2.623      0.013     142.145     982.606
ar.L1.y        1.4218      0.146      9.712      0.000       1.135       1.709
ar.L2.y       -0.4938      0.157     -3.149      0.003      -0.801      -0.186
                                    Roots
=============================================================================
                 Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           1.2213           +0.0000j            1.2213            0.0000
AR.2           1.6584           +0.0000j            1.6584            0.0000
-----------------------------------------------------------------------------
```



As we can see the model describes the actual values rather good.

5 Forecast

5.1 Multiple regression

$current\_price = exp(-11.5109 + 3.9755 * log(ppiaco) - 0.6026 * log(will5000ind))$

Assuming that in 2015 ppiaco was 190.4 and market index (will5000ind) was 90.66, we can find gold price in 2015 = 766 USD per troy ounce.

5.2 ARIMA

The designed model gives us price value **1143.79641724** for the next year.

I must say that the difference between two models is considerable. But as R^2 says ARIMA result is a little bit more reliable.