International Journal of Applied Mathematics

Volume 26 No. 4 2013, 425-431
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
doi: http://dx.doi.org/10.12732/ijam.v26i4.2

INFORMATION WORTH OF OBSERVATIONS IN PREDICTION OF MONTHLY AVERAGE OF GOLD PRICE

Mostafa Pouralizadeh[1 §], Mojgan Pouralizadeh[2]
[1,2]Department of Mathematics
Azad University of Lahijan
Lahijan, Postal Code: 44169-39515, IRAN

Abstract: In this study, we try to quantify the information worth of observation in the monthly average of gold price for prediction of future outcome. Furthermore, with the fitness of an ARMA model to the data under the condition of contamination, we show how to improve the precision of prediction.

AMS Subject Classification: 28D20, 37M10
Key Words: (ACF) Auto Correlation Function, (PACF) Partial Auto Correlation Function, (ARMA) Auto Regressive Moving Average, stationary time series

1. Introduction

The aim of information worth of observation is that, how much sufficient information there is in predictive observations for an accurate prediction. According to this study, first of all quantitative value is determined for the information worth of observations, then in prediction of a realization in a time series, we can use observations which are more worthy than others. A quantitative value for the information worth of observation in prediction of a normal stationary time series realization, is detailed in Pourahmadi [1].

In this study we try to use these quantitative measures, worth in prediction

Received: May 18, 2013

§Correspondence author

of monthly average of 1 gram,18 assay gold price in Iran market. In the second section by implementing Box and Jenkins [3] method, common and corresponded ARMA models are fitted to the second lag difference of data. In the third section using information theory and entropy, information worth of observations is computed, and in the forth section, under condition of contamination and irregularity in predictive observations, the results are analyzed and the best model is chosen.

2. Fitting ARMA(p,q) to the Data

In this section, we implement some statistical methods to fit three adequate models to the data. Since the economic crisis has been increasing in all over the world, it affects on all aspects of marketing and goods prices, so we guess that gold price might be involved by contamination and irregularity. To detect contamination of data we miss X_{40} out of data and implement time series analysis under previous 39 data to predict X_{40}. So contamination and side effects can be detected by computing the information worth of observation in prediction of X_{40}.

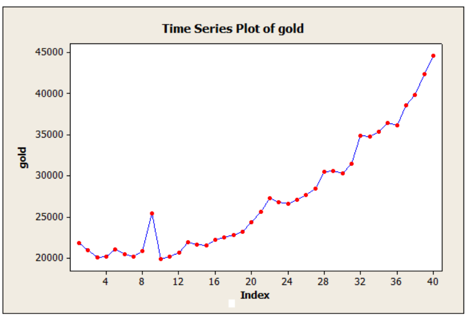

Figure 1: Time series plot of data

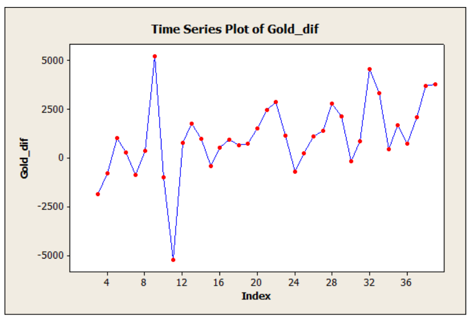

Figure 2: Time series plot of fitted data

Looking at Figure 1, it can easily be inferred that there is an increasing trend in data; in addition, normality test has shown that observations are not from normal population. Hence, we use second lag difference of data to eliminate increasing trend from data and to make them normal and stationary. Figure 2 indicates second lag difference of data; moreover, normal probability plot, auto correlation function and partial auto correlation of data are indicated by the following figures, so the best adequate models could be fitted to the data, working toward creating an accurate prediction.

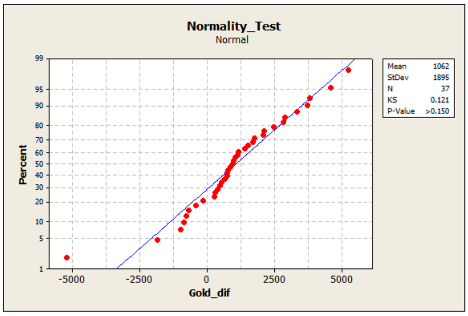

Figure 3: Normal probability plot of fitted data

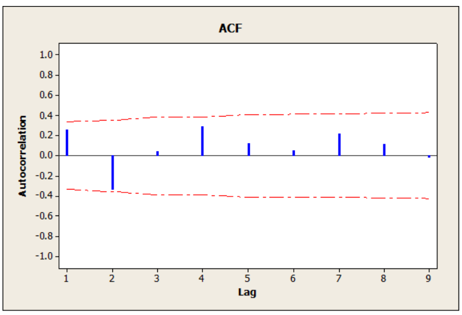

Figure 4: ACF of fitted data

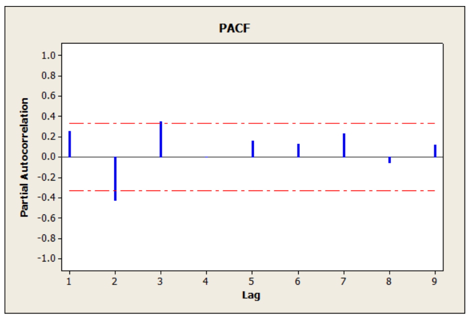

Figure 5: PACF of fitted data

The Kolomogrov and Smirnov test and the normal probability plot in Figure 3 shows that the fitted data are from a normal population (P-Value$>$ 0.150). From Figure 5, it is clear that PACF of data shows a maximum absolute value in lag 2, so it can be realized that AR(2) is an adequate model for the data. Also according to the treatment of ACF, we fit ARMA(2,2) to the data as a competitive model for AR(2); furthermore, by the reason, we will point out in the third section, AR(5) is fitted to the data too. This statistical approach is based on the Box and Jenkins method of fitting [3], and using the software Minitab, Version 15, these three models are fitted to the second lag difference of observations, and the forecast of X_{40} is computed for each of the three competitive models. The estimations of the model parameter and the forecast of X_{40} are defined by the following equations:

$$\begin{cases} ARMA(2,2) : X_t = 0.2599X_{t-1} - 0.1331X_{t-2} + Z_t + \\ \qquad 1.7384Z_{t-1} - 0.8237Z_{t-2} \\ \widehat{X}_{40} = 42304.9 \end{cases}, \tag{1}$$

$$\begin{cases} AR(2) : X_t = -0.7933X_{t-1} - 0.4112X_{t-2} \\ \widehat{X}_{40} = 44342.6 \end{cases}, \tag{2}$$

$$\begin{cases} AR(5) : Xt = -1.2405X_{t-1} - 1.2615X_{t-2} - 1.0840X_{t-3} - \\ \qquad 0.5798X_{t-4} - 0.2129X_{t-5} \\ \widehat{X}_{40} = 42472.6 \end{cases}. \tag{3}$$

3. Worth of Observation in Prediction

In this section we compute the forecast of X_{40} based on the previous five observations, and that is why we fitted AR(5) to the data. The general form of the auto regressive-moving average models is:

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}. \tag{4}$$

Here ϕ_i's are the auto regressive coefficients and θ_i's are the moving average coefficients in this model.

A quantitative value of information worth existed in X_{-m} for prediction of X_0 in stationary normal time series is computed by the following formula, [1]:

$$w = \frac{{\pi_m}^2}{1 + \sum_{i=1}^{n} {\pi_i}^2}. \tag{5}$$

Here m is the time difference between X_{-m} and X_0 and the π_i's are earned by the following power series:

$$\frac{\phi(z)}{\theta(z)} = \sum_{i=0}^{\infty} \pi_i z^i, \qquad \pi_0 = 1, \tag{6}$$

where

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$$

and

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q.$$

First of all in each of the three models we compute the π_i's based on 6, then using 5 we calculate worth of each of the previous five observations of X_{40} for $m = 1, 2, \cdots, 5$ to compare three selected models with each other. In the following table, the quantitative worth existing in each of the five previous observations of X_{40} is shown:

m	1	2	3	4	5
AR(2)	0.3862	0.1955	0	0	0
AR(5)	0.6061	0.0293	0.0071	0.1149	0.016
ARMA(2,2)	0.7997	0.7871	0.7798	0.7791	0.7787

Table 1: Worth of each of the previous five observation in prediction of X_{40}

To create a profound insight for the readers, Figure 6 shows the worth of each single observation in prediction of X_{40}.

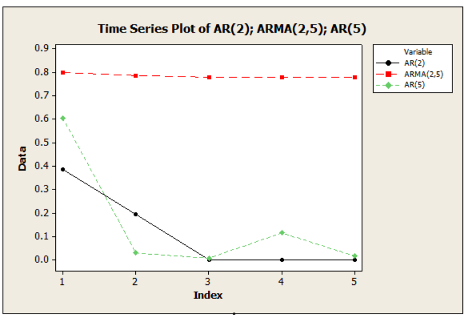

Figure 6 : Worth of the previous five observations of fortieth step

It can be inferred from Figure 6 that, in all steps ARMA(2,2) is more worthy than others in prediction of X_{40}, so if in these steps the data are not exposed by destructive phenomena, ARMA(2,2) would be the best model for prediction of X_{40} among three competitive models. Also AR(2) in most steps gain little worth in prediction of X_{40}, in comparison with its two competitors. Therefore, since AR(2) is less worthy than others in prediction of X_{40}, we can choose it as the best model in prediction if there are contaminations and destructive factors in each of the 5 previous observation of X_{40}.

4. Discussion

The best way to decrease the side effect of natural phenomena is that, if each of the previous five observations of X_{40} is exposed by destructive factors, we choose the model which is less worthy than others. In the other words, we can rank the three competitive models in order to accuracy in prediction; in addition, if there is not any contamination in data we can rely on the models which are more worthy than others and put them in priority.

Figure 6 shows that for the observations X_{35}, X_{36}, X_{37} and X_{39} the model AR(2) is less worthy than its competitors, and ARMA(2,1) is the most worthy model for prediction in each of the previous five observations of X_{40}; moreover, to predict X_{40} based on X_{38}, its clear that ARMA(2,2) gain the least worth in comparison with the two other models. Hence it can be realized that, in most steps of the previous five observations of X_{40}, the model AR(2) is the least worthy model in prediction of X_{40}, but the forecast of X_{40} based on AR(2) ($\widehat{X}_{40} = 44342.6$) is the closest estimation to the real value of X_{40} ($X_{40} = 44577$). Therefore, the model with the smallest worth in prediction, states the best forecast for the real value of X_{40}. On the other hands, ARMA(2,2) is the most worthy model in prediction of X_{40} based on each of the previous five observations of X_{40}, but its forecast for X_{40} (42304.9) is the farthest estimation to the real value of X_{40} among three competitive models. So we can conclude that the previous five observations of X_{40} are influenced by contamination, and to predict gold price in Tir 1390 ($\widehat{X}_{41}$) based on the previous six observations, we can rely on AR(2), and the next priorities are filled by AR(5) and ARMA(2,2), respectively.

Hence, the best forecast for monthly average of gold price in Tir, 1390 is 46521.7 which is based on $AR(2)$.

5. Conclusions

To sum up, this practical method can be implemented under conditions of contamination in some fields like meteorology, economy and marketing, so it helps decrease deviations in interpretation of results, and provides a great deal of precision in prediction of future realizations of a time series; furthermore, this statistical approach can be used for prediction whether observations are contaminated or not. In practical conditions, if in a stationary normal time series, there are some reports of side effects, (probable economic and political crisis, flood, fire and etc.) the least worthy model among several adequate models is the best one in prediction. In this study we investigated 1 gram 18 assay gold price in Iran market, but this statistical method can be executed in other aspects of economic to enhance the precision of forecast of a time series realizations.

References

[1] M. Pour Ahmadi, E. Soofi, Prediction variance and information worth of observation in time series, *J. Time. Ser. Anal.*, **21** (2000), 413-434.

[2] K. Choy, Outlier detection for stationary time series, *J. Stat. Plann. Infer.*, **99** (2001), 111-127.

[3] G.E.P Box, G.M. Jenkins, G.C. Reinsel, *Time Series Analysis, Forecasting and Control*, Prentice Hall, New Jersey (1994).

[4] T. Cover, S. Thomas, *Elements of Information Theory*, Wiley, N. York (1990).

[5] P.J. Brockwell, R.A. Davis, *Time Series: Theory and Method*, Springer, N. York (1990).