A study on mathematical methods for predicting accuracy of crude oil futures prices by multi grey Markov model

S. Indrakala¹*

Abstract
In the world economy, Crude oil is one of the most important fuel material and and its price affects the price of many other commodities, including gasoline and natural gas. However, the flow effect of crude oil prices also impacts the price of stocks, bonds, and currencies around the globe. In this situation, industries, governments and individuals pay special attention in crude oil price Prediction. Even if, several mathematical models have been established for predicting oil prices, it has more difficulties to forecast due to the high irregularity of oil prices. In this paper, we propose a new approach for crude oil price prediction based on Multivariate Grey Model with Markov Model and present the greater precision compared to the traditional Multivariate Grey Model.

Keywords
Crude oil, economic growth, Multivariate Grey Model.

1 Department of Mathematics, Kunthavai Nacchayir Government Arts College for Women, Thanjavur-613007, Tamil Nadu, India.
*Corresponding author: s.indirakala@yahoo.com

Article History: Received 19 December 2020; Accepted 19 February 2021 ©2021 MJM.

1 Introduction
Crude oil is the leading component of energy that influences the growth for industrialized and developing countries and also the economic development of a global economy. The demand for crude oil will continue to increase, although its pace of growth is expected to slow gradually, according to the Indian Petroleum energy outlook 2017. Owing to the significance of crude oil, a lot of investors, governments and enterprises pay much attention to the trend of crude oil prices. Due to the factors such as assumption activities, supply and demand, geopolitical clashes and wars can greatly produce effects on the prices of crude oil. Therefore, it is a challenging task to forecast the crude oil prices accurately. Various models have emerged to try to forecast the crude oil prices as accurately as possible in recent years. Generally, the prediction models has been classified into two types. The first type is statistical and econometric models which includes vector autoregressive (VAR) models [4], the random walk model (RWM) [8, 16], the autoregressive integrated moving average (ARIMA) [1,2] and generalized autoregressive conditional heteroskedasticity (GARCH) family models [15,3] and the second type is an artificial intelligence (AI) models [12]. For example, Huntington (1994) used a sophisticated econometric model with a variety of impact factors to predict crude oil price. Baumeister and Kilian (2014) established that VAR models for forecasting crude oil prices at short forecast horizons with achieve good accuracy [4]. Hou and Suardi (2012) implemented a nonparametric GARCH model to predict the return volatility in oil price. Panopoulou and Pantelidis established two and three-state regime switching models to forecast crude oil prices. Wang and Wu forecasted the volatility of crude oil prices using multivariate GARCH-class models, and the results indicated that the multivariate models showed better performance than univariate models [17]. Some other researchers focused on a multivariate analysis of crude oil prices.

However, all the statistical and econometric models have the limitation that they are built on the assumption that the time series of crude oil prices has the characteristics of lin-
A study on mathematical methods for predicting accuracy of crude oil futures prices by multi grey Markov model

2. Modeling mechanism of Multivariate Grey Markov Model

Suppose the non-negative original sequence is

$$c_i^{(0)}(j) = [c_i^{(0)}(1), \ldots, c_i^{(0)}(j), \ldots, c_i^{(0)}(m)]$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$, where $n$ is the number of variables and $m$ is the sequence length of each variable, then

$$c_i^{(1)} = [c_i^{(1)}(1), \ldots, c_i^{(1)}(j), \ldots, c_i^{(1)}(m)]$$

is the first-order accumulated generation of $c_i^{(0)}$ and it is denoted by $1$-AGO as follows

$$c_i^{(1)}(j) = \sum_{j=1}^{k} c_i^{(0)}(j) \quad (j = 1, 2, \ldots, m) \quad (2.1)$$

Let the corresponding Multivariate raw data matrix be

$$\mathcal{C}^{(0)} = \begin{bmatrix} c_1^{(0)}(1) & c_1^{(0)}(2) & \cdots & c_1^{(0)}(m) \\ c_2^{(0)}(1) & c_2^{(0)}(2) & \cdots & c_2^{(0)}(m) \\ \vdots & \vdots & \ddots & \vdots \\ c_n^{(0)}(1) & c_n^{(0)}(2) & \cdots & c_n^{(0)}(m) \end{bmatrix}$$

and

$$\mathcal{C}^{(1)} = \begin{bmatrix} c_1^{(1)}(1) & c_1^{(1)}(2) & \cdots & c_1^{(1)}(m) \\ c_2^{(1)}(1) & c_2^{(1)}(2) & \cdots & c_2^{(1)}(m) \\ \vdots & \vdots & \ddots & \vdots \\ c_n^{(1)}(1) & c_n^{(1)}(2) & \cdots & c_n^{(1)}(m) \end{bmatrix} \quad (2.2)$$

Then equation (2.3) can be written as

$$\frac{d\mathcal{C}^{(1)}}{dt} = A\mathcal{C}^{(1)} + B \quad (2.4)$$

The first component $c_i^{(1)}(1)$ of the sequence $c_i^{(1)}(j) \quad (j = 1, 2, \ldots, m)$ is taken as an initial condition of the grey differential equation and the continuous time response of Eq. (4) is as

$$\mathcal{C}^{(1)}(t) = e^{At}\mathcal{C}^{(1)}(1) + A^{-1}(e^{At} - I)B \quad (2.5)$$

Where,

$$e^{At} = I + \sum_{k=1}^{\infty} \frac{A^k}{k!}t^k$$

Here $I$ is an unit matrix. With known matrix $A$ and time $t$, we can easily calculate exponent matrix $e^{At}$ without appearing singular matrix with function. $\mathcal{C}^{(0)} = AP^{(1)} + B$ can be found through the difference grey derivative in $[k - 1, k]$. Let us take $P^{(1)}(k) = 0.5 \ast \left( c_i^{(1)}(k) + c_i^{(1)}(k - 1) \right)$, then the following equation can be obtained as

$$c_i^{(0)}(k) = \sum_{j=1}^{n} \frac{a_{ij}}{2} \left( c_i^{(1)}(k) + c_i^{(1)}(k - 1) \right) + b_i \quad (2.6)$$

Where $i = 1, 2, \ldots, n$ and $k = 2, 3, \ldots, m$

Suppose $a_i = (a_{i1}, a_{i2}, \ldots, a_{in}, b_i)^T, (i = 1, 2, \ldots, n)$ then the identified $\hat{a}_i$ of $a_i$ can be obtained through the least square method as follows

$$\hat{a}_i = (\hat{a}_{i1}, \hat{a}_{i2}, \ldots, \hat{a}_{in}, \hat{b}_i)^T = (P^T P)^{-1}P^T Q_i \quad (2.7)$$

where $i = 1, 2, \ldots, n$. 

The Multivariate Grey model is the first-order differential equation with $n$ variables which is given below

$$\begin{cases}
\frac{dc_i^{(1)}}{dt} &= a_{11}c_i^{(1)} + a_{12}c_2^{(1)} + \cdots + a_{1n}c_n^{(1)} + b_1 \\
\frac{dc_2^{(1)}}{dt} &= a_{21}c_1^{(1)} + a_{22}c_2^{(1)} + \cdots + a_{2n}c_n^{(1)} + b_2 \\
&\vdots \\
\frac{dc_n^{(1)}}{dt} &= a_{n1}c_1^{(1)} + a_{n2}c_2^{(1)} + \cdots + a_{nn}c_n^{(1)} + b_n 
\end{cases} \quad (2.3)$$

Here,

$$\mathcal{C}^{(0)}(k) = \begin{bmatrix} c_1^{(0)}(k), c_2^{(0)}(k), \ldots, c_n^{(0)}(k) \end{bmatrix}^T$$

$$\mathcal{C}^{(1)}(k) = \begin{bmatrix} c_1^{(1)}(k), c_2^{(1)}(k), \ldots, c_n^{(1)}(k) \end{bmatrix}^T$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\
 a_{21} & a_{22} & \cdots & a_{2n} \\
 \vdots & \vdots & \ddots & \vdots \\
 a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \quad B = \begin{bmatrix} b_1 \\
b_2 \\
\vdots \\
b_n \end{bmatrix}$$

Owing to the high nonlinearity and non-stationarity, it is usually inflexible for these models to be directly applied to crude oil price forecasting to achieve satisfactory results. Therefore, the AI models have concerned increasing curiosity for crude oil price forecasting because they can capture the intrinsic features of nonlinearity and non-stationarity that exist widely in crude oil prices. As for AI models, the most popular ones include the artificial neural network (ANN) and support vector regression (SVR). Deng [5] in 1982 proposed the grey forecasting forecasting model which is superior to deal with small data set and poor information [6], and has been widely applied in many forecasting research fields. GM (1,1) is the most commonly grey model, which reveals the inherent development law by the first-order differential equation with single variable [1]. GM(1,1) with single variable was extended to the multivariate grey model MGM (1,n) [2]. MGM (1, n) is neither the simple combination of GM(1,n) nor GM(1,n) that establishing only a first-order differential equation with $n$ variables. In MGM (1, n), $n$ differential equations with $n$ variables are established and solved. The proposed hybrid forecasting technique can successfully handle the information of available historical data and increase the forecasting precision. Markov method predicts the future states of the system by transition probability matrix between different system states. The combined model consists of three stages. In the first stage, Multi Grey model is developed to fit the historical statistical data and in the next stage residual errors will be classified from the output using Markov chain and the last stage is to predict future values that might be supplied by the hybrid Grey model.
and $Q_t = \left( c_i^{(0)}(2), c_i^{(0)}(3), \ldots, c_i^{(0)}(m) \right)^T$.

Let $p$ and $\hat{p}$ be the original sequence of $n$ variables and the predicted residual error value of the Multi-Grey models respectively. Based on the predicted sequence $\hat{p}$, a residual error sequence is defined as follows

$$\delta = \{\delta(2), \delta(3), \ldots, \delta(k), \ldots, \delta(n)\} \quad (2.8)$$

where $\delta(k) = \frac{p(k)}{p(k)}$ with $k = 2, 3, 4, \ldots, n$.

Next, we create Markov state transition matrices from the predicted residual error value of the Multi-Grey models. Then, the states are defined as follows

State 1 = \left[ \min(\delta(k)), \min(\delta(k)) + \frac{\max(\delta(k)) - \min(\delta(k))}{m} \right]

State $i = \left[ \min(\delta(k)) + \frac{\max(\delta(k)) - \min(\delta(k))}{m} \cdot (i - 1), \min(\delta(k)) + \frac{\max(\delta(k)) - \min(\delta(k))}{m} \cdot i \right]$, $i = 1, 2, 3, \ldots, m - 1$

State $m = \left[ \min(\delta(k)) + \frac{\max(\delta(k)) - \min(\delta(k))}{m} \cdot (m - 1), \max(\delta(k)) \right]$

Since probabilities are nonnegative and the method must sort a transition into some state, we have

$$P_{ij} \geq 0, \quad i, j \geq 0, \quad \sum_{j=1}^{\infty} P_{ij} = 1, \quad i = 0, 1, 2, \ldots$$

The one step transition probability of state is $P_{ij} = \frac{Q_{ij}}{Q_t}$, $j = 1, 2, \ldots, r$ so that

$$P_{ij} = \begin{bmatrix}
P_{i1} & P_{i2} & \cdots & P_{ij} \\
P_{i1} & P_{i2} & \cdots & P_{ij} \\
\vdots & \vdots & \ddots & \vdots \\
P_{i1} & P_{i2} & \cdots & P_{ij}
\end{bmatrix}$$

$P_{ij}^{(m)}$ is the probability of transition from state $i$ to $j$ by $m$ step.

$Q_{ij}^{(m)}$ is the transition time from state $i$ to $j$ by $m$ steps and $Q_t$ is the number of data belonging to the $i$ th state. Because the transition for the last $m$ entries of the series is indefinite, $Q_t$ should be counted by the first as $n - m$ entries. After determining the state and the probability matrix, we predict the future data using the following formula

$$\hat{\delta}(k) = \delta^{(0)}(k) + (\text{Probability of state} \times 100)$$

3. Application of Multivariate Grey Markov for oil price prediction

In this section, Crude oil spot price series were chosen as the experimental data. Besides being one of the most actively traded commodities, the price of crude oil is extremely sensitive to geopolitical and weather events. In fact, the world crude oil market is all about investor expectation of supply and demand and oil prices are very volatile and greatly influenced by consumer and investor sentiment. Here we take the daily trend price of crude oil from December 12, 2011 to December 06, 2019 and the graphical representation of the price of WTI crude oil is given in Figure 1.

![Figure 1. Trend line of WTI Crude oil Price](image)

The parameters of grey MGM (1, n) model are as follows:

| S.No | Value         | A          | B          |
|------|---------------|------------|------------|
| 1    | Price         | -4.06E-01  | 4353.169063|
| 2    | Open          | -4.05E-01  | 4357.7869  |
| 3    | High          | -4.11E-01  | 4.40E+03   |
| 4    | Low           | -4.01E-01  | 4303.528   |
| 5    | Volume        | -1.59E-02  | 149.5199194|

From table 1, we obtain

$\hat{\gamma}(1)(t) = 4212e^{-4.06E-01t} - 1.07E + 04 \cdot \left( e^{-4.06E-01t} - 1 \right) \quad (\text{prices})$

$\hat{\gamma}(1)(t) = 4154e^{-4.05E-01t} - 1.0731.4916 \cdot \left( e^{-4.05E-01t} - 1 \right) \quad (\text{open})$

$\hat{\gamma}(1)(t) = 4265e^{-4.11E-01t} - 1.07E + 04 \cdot \left( e^{-4.11E-01t} - 1 \right) \quad (\text{high})$

$\hat{\gamma}(1)(t) = 4115e^{-4.01E-01t} - 1.0735.2 \cdot \left( e^{-4.01E-01t} - 1 \right) \quad (\text{low})$

$\hat{\gamma}(1)(t) = 304.45e^{-1.59E-02t} - 9375.37509 \cdot \left( e^{-1.59E-02t} - 1 \right) \quad (\text{Volume})$

The forecasted values are calculated from MGM(1, n) model using the above equations and the corresponding residual error sequences are calculated by equation (7). Then, the state intervals of residual error sequences are classified as follows.
The initial distribution of stock price values can be derived as the study state distribution can be calculated after m-step therefore useful for both marketing and investing. Generally, the study state distribution can be calculated after m-step transition from the initial phase is \( S^{(m)} = S^{(0)} \cdot P^{(m)} \). After solving the Equation, we obtain the steady state probabilities as follows:

The forecasted values from 2011 to 2019 by \( \text{MGM}(1, n) \) and Markov MGM \((1, n)\) are compared and listed in the following tables and figures

The WTI Crude oil Prices are categorized into four possible states according to the prices. State intervals of residual error of WTI Crude oil Prices are labeled in Table 2. Such labels are therefore useful for both marketing and investing. Generally, the study state distribution can be calculated after m-step transition from the initial phase is \( S^{(m)} = S^{(0)} \cdot P^{(m)} \). After solving the Equation, we obtain the steady state probabilities as follows:

The forecasted values from 2011 to 2019 by \( \text{MGM}(1, n) \) and Markov MGM \((1, n)\) are compared and listed in the following tables and figures

From the table (2), we predict the future stock price values. The initial distribution of stock price values can be derived as

\[
S^{(0)}_{\text{Price}} = (0.089497, 0.402274, 0.461724, 0.046507)
\]

\[
S^{(0)}_{\text{Open}} = (0.111929, 0.3298019, 0.375403, 0.182865)
\]

\[
S^{(0)}_{\text{High}} = (0.1497, 0.2151, 0.3463841, 0.288807)
\]

\[
S^{(0)}_{\text{Low}} = (0.369415, 0.069553, 0.3298019, 0.230796)
\]

\[
S^{(0)}_{\text{Volume}} = (0.301704, 0.2266236, 0.27268539, 0.19898664)
\]

The one-step transition probability matrices are as follows

\[
P_{\text{Price}} = \begin{bmatrix}
0.300518 & 0.606218 & 0.07772 & 0.015544 \\
0.152174 & 0.360412 & 0.448513 & 0.038902 \\
0.001994 & 0.434696 & 0.500499 & 0.062812 \\
0.01 & 0.05 & 0.93 & 0.01
\end{bmatrix}
\]

\[
P_{\text{Open}} = \begin{bmatrix}
0.3868501 & 0.25535168 & 0.1957186 & 0.162079 \\
0.25 & 0.1930894 & 0.4126016 & 0.1443089 \\
0.25675675 & 0.375 & 0.217905 & 0.1503378 \\
0.2037037 & 0.259259 & 0.224537 & 0.3125
\end{bmatrix}
\]

\[
P_{\text{High}} = \begin{bmatrix}
0.3661538 & 0.15076923 & 0.261538 & 0.16 \\
0.334047 & 0.2098501 & 0.24625267 & 0.209850 \\
0.29654255 & 0.1702127 & 0.190159 & 0.343085 \\
0.276358 & 0.140575 & 0.380192 & 0.2028754
\end{bmatrix}
\]

\[
P_{\text{Low}} = \begin{bmatrix}
0.291764 & 0.223529 & 0.190588 & 0.294117 \\
0.381703 & 0.2302839 & 0.223975 & 0.164038 \\
0.383954 & 0.265043 & 0.133238 & 0.217765 \\
0.302739 & 0.216438 & 0.1726027 & 0.3082191
\end{bmatrix}
\]

\[
P_{\text{Volume}} = \begin{bmatrix}
0.1920123 & 0.322581 & 0.33978 & 0.145929 \\
0.2540650 & 0.3069105 & 0.27439 & 0.1646341 \\
0.544715 & 0.264228 & 0.4451291 & 0.52439 \\
0.3134328 & 0.1582089 & 0.3462687 & 0.182089
\end{bmatrix}
\]
A study on mathematical methods for predicting accuracy of crude oil futures prices by multi grey Markov model — 625/626

4. Conclusion

Forecasting crude oil prices is a very challenging problem due to the high volatility of oil prices. In this paper Multivariate Grey model and Markov Model is established to predict the trend of crude oil price values. Based on the whole modeling and prediction, we can conclude that our model is quite easy to handle and capture the main feature. The crude oil price can suffer sudden increase and decrease. The experiment results show that the Multivariate Grey Model with Markov Model is the most efficient grey forecasting model because of its better forecasting accuracy.
A study on mathematical methods for predicting accuracy of crude oil futures prices by multi grey Markov model —

References

[1] Abledu, G.K.; Agbodah, K. Stochastic Forecasting and Modelling of Volatility of Oil Prices in Ghana using ARIMA Time series model. Eur. J. Bus. Manag. 4(2012), 122–131.

[2] Ahmed, R.A.; Shabri, A.B. Daily crude oil price forecasting model using arima, generalized autoregressive conditional heteroscedastic and support vector machines. Am. J. Appl. Sci. 11(2014), 425.

[3] Arouri, M.E.H.; Lahiani, A.; Lévy, A.; Nguyen, D.K. Forecasting the conditional volatility of oil spot and futures prices with structural breaks and long memory models. Energy Econ. 34(2012), 283–293.

[4] Baumeister, C.; Kilian, L. Real-time forecasts of the real price of oil. J. Bus. Econ. Stat. 30(2012), 326–336.

[5] Deng, J.L. Control problems of grey systems. Syst. Control Lett. 1(1982), 288–294.

[6] Ding, S.; Hipel, K.W.; Dang, Y. Forecasting China’s electricity consumption using a new grey prediction model. Energy 2018, 149, 314–328.

[7] Kang, Sang Hoon, Sang-Mok Kang, and Seong-Min Yoon. “Forecasting volatility of crude oil markets.” Energy Economics, 31(1)(2009), 119–125.

[8] Hooper, V.J.; Ng, K.; Reeves, J.J. Quarterly beta forecasting: An evaluation. Int. J. Forecast. 24(2008), 480–489.

[9] Huntington Hillard G. Oil Price Forecasting in the 1980s: What Went Wrong? The Energy Journal, 15(2)(1994), 1-22.

[10] Hou Aijun, Suardi Sandy. A nonparametric GARCH model of crude oil price return volatility. Energy Economics, 34(2)(2012), 618–626.

[11] Moshiri, Saeed, and Faezeh Foroutan. “Forecasting nonlinear crude oil futures prices.” The Energy Journal, (2006), 81-95.

[12] Li, T.; Zhou, M.; Guo, C.; Luo, M.; Wu, J.; Pan, F.; Tao, Q.; He, T. Forecasting Crude Oil Price Using EEMD and RVM with Adaptive PSO-Based Kernels. Energies, 9(2016), 1014.

[13] Y.X. Luo, W.Y. Xiao, New Information Grey Multivariable Optimizing Model NMGM(1,n,q,r) for the Relationship of Cost and Variability, 2009 International Conference on Intelligent Computation Technology and Automation (ICICTA 2009), 120-123. 2009

[14] Mirmirani, S.; Li, H.C. A comparison of VAR and neural networks with genetic algorithm in forecasting price of oil. Adv. Econom. 19(2004), 203–223.

[15] Morana, C. A semiparametric approach to short-term oil price forecasting. Energy Econ. 23(2001), 325–338.

[16] Murat, A.; Tokat, E. Forecasting oil price movements with crack spread futures. Energy Econ. 31(2009), 85–90.

[17] Wang, Y.; Wu, C. Forecasting energy market volatility using GARCH models: Can multivariate models beat univariate models? Energy Econ. 34(2012), 2167–2181.

[18] J. Zhai, J.M. Sheng, Grey Model and its Application, Systems Engineering-Theory & Practice. 17(5)(1997),