Al-Ghary Journal of Economic and Administrative Sciences Vol. 20 (No.2) 2024 PP. 42-64

مجلة الغري للعلوم الاقتصادية والادارية

مجلد (20) عدد (2) 2024

Forecasting Price of Crude Oil Using the Weight Markov Chain (WM-CM) and ARIMA Model Techniques

Assistant Prof. Dr.Mohammad Mahmood Faqe Hussein
University of Sulaimani/College of Admin&Eco.
Mohammad.faqe@univsul.edu.iq

Abstract

Crude oil prices play a pivotal role in the global economy, influencing everything from energy expenses to inflation rates. Accurately predicting these prices is crucial for governments, businesses, and investors. The aim for this study is utilizing Weighted Markov Chain (WM-CM) and ARIMA (Autoregressive Integrated Moving Average) model techniques for crude oil price forecasting is to ensure precise estimations of forthcoming crude oil prices. These methods present distinct strategies for scrutinizing historical price data and projecting future trends. By integrating these approaches, there's a prospect of augmenting the precision and dependability of forecasts. We collected data from the period of January 1, 2000, to September 29, 2023, for crude oil prices, spanning 285 months, from the website "investing.com". We compared two models using metrics mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and AIC. Our study led us to conclude that the ARIMA (2,0,0) model outperforms the weighted Markov Chain (WMC) model. This conclusion is supported by lower values for mean absolute error, root mean square error, mean absolute percentage error, and AIC. Additionally, the predictive performance of the chosen model for crude oil prices showed an increasing trend for the months of October and November, followed by a decline for the months of December, January, February, March, April, May, June, July, August, and September.

Keywords: Forecasting, Weighted Markov Chain, ARIMA, MAPE, MAE, AIC

التنبؤ بسعر النفط الخام باستخدام نموذج سلسلة ماركوف الموزون (WM-CM)وتقنيات النموذج ARIMA

الأستاذ المساعد الدكتور محمد محمود فقى حسين
جامعة السليمانية / كلية الإدارة والاقتصاد

المستخلص

تلعب أسعار النفط الخام دورًا محوريًا في الاقتصاد العالمي، حيث تؤثر على كل شيء بدءًا من نفقات الطاقة وحتى معدلات التضخم. يعد التنبؤ الدقيق بهذه الأسعار أمرًا بالغ الأهمية للحكومات والشركات والمستثمرين. الهدف من هذه الدراسة هو استخدام تقنيات نموذج سلسلة ماركوف الموزون (WM-CM) وARIMA (المتوسط المتحرك المتكامل الذاتي) للتنبؤ بأسعار النفط الخام لضمان تقديرات دقيقة لأسعار النفط الخام القادمة. تقدم هذه الأساليب استراتيجيات متميزة لفحص بيانات الأسعار التاريخية وتوقع الاتجاهات المستقبلية. ومن خلال دمج هذه الأساليب، هناك احتمال لزيادة دقة وموثوقية التنبؤات. قمنا بجمع بيانات من الفترة من 1 يناير 2000 إلى 29 سبتمبر 2023، لأسعار النفط الخام، على مدى 285 شهرًا، من موقع "Investing.com". وقمنا بمقارنة نموذجين باستخدام مقاييس متوسط الخطأ المطلق (MAE)، الجذر متوسط مربع الخطأ (RMSE)، ومتوسط النسبة المئوية للخطأ المطلق (MAPE)، وAIC. قادتنا دراستنا إلى استنتاج أن نموذج ARIMA (2،0،0) يتفوق على نموذج سلسلة ماركوف الموزون (WM-CM). ويدعم هذا الاستنتاج قيم أقل لمتوسط الخطأ المطلق، وجذر متوسط مربع الخطأ، ومتوسط النسبة المئوية للخطأ المطلق، وAIC. بالإضافة إلى ذلك، أظهر الأداء التنبؤي للنموذج المختار لأسعار النفط الخام اتجاهاً تصاعدياً لشهري أكتوبر ونوفمبر، يليه انخفاض للأشهر ديسمبر، يناير، فبراير، مارس، أبريل، مايو، يونيو، يوليو، أغسطس. ، وسبتمبر.

الكلمات المفتاحية: التنبؤ، سلسلة ماركوف المرجحة، ARIMA، MAPE، MAE، AIC

1 Introduction

Crude oil, commonly known as "black gold," is an essential commodity with a pivotal role in the global economy. It serves as a primary energy source for transportation, manufacturing, and various other industries. The determination of crude oil prices involves a complex interaction of factors such as supply, demand, geopolitical events, market sentiment, and economic indicators. Understanding the intricacies of crude oil price dynamics is crucial for policymakers, investors, and industry stakeholders.

Derived from underground reservoirs, crude oil is a natural fossil fuel comprising hydrocarbons of varying compositions. Following extraction, it undergoes refining processes to yield products like gasoline, diesel, jet fuel, and petrochemicals.

Given its widespread use as an energy source, the price of crude oil holds paramount importance. Fluctuations in crude oil prices carry significant implications for consumer spending, inflation rates, corporate profitability, and government revenues.

2. Literature review

Zhang and Wang (2017) proposed a novel approach that integrates WMC and ARIMA models for forecasting crude oil prices. They demonstrated that combining the transitional probabilities captured by WMC with the time series analysis provided by ARIMA improves forecast accuracy. Aloui et al. (2017) evaluated the performance of ARIMA models in forecasting crude oil prices, considering various factors such as volatility clustering and seasonality. They found that ARIMA models can effectively capture short-term fluctuations in crude oil prices. Kang and Ratti (2021) conducted a comprehensive analysis of ARIMA models, focusing on different specifications and parameter tuning techniques. Their study highlighted the importance of selecting appropriate lag orders and differencing parameters for accurate forecasts. Liu et al. (2020) proposed an adaptive WMC model that dynamically adjusts the weights assigned to transitional probabilities based on the volatility of crude oil prices. Their approach showed improved accuracy compared to traditional WMC models.

3 Methodology

3.1 Markov Chain Prediction Model Principle

Assuming that for any instance, when the process $\{X(t)\}$ is in state i, the probability that the process $X(t)\}$ will transition to state j at the next time step is a constant value denoted as P_{ij} . A Markov chain is a specific type of stochastic process, and its definition is as follows: For any state of the process $\{X(t)\}$ in the past, represented as a nonnegative integer, the following formula holds:[5]

$$P_{ij} = P\{X_{n+1} = j|X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\}$$
$$= P\{X_{n+1} = j|X_n = i_n\} \text{ where } i \geq 0 \text{ and } j \geq 0 \quad (1)$$

The Markov process is a random process.[10]

3.2 Procedure of weighted Markov chain:

1. Estimate the average value, symbolized as $\bar{x}$, and the Std., using old data. Categorize the states based on how they relate to the levels defined by the average and Std. [7],[8],[10]

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} \quad (2)$$

$$\text{Std.} = \text{SQRT}\left[\frac{\sum_{i=}^{n}(x_i - \bar{x})^2}{n-1}\right] \quad (3)$$

2. Classify the old data into m unique states by utilizing the average $\bar{x}$ and std. calculated from the old dataset.

Table (1): Represent an example of m states

States	Levels
One	$x < \bar{x} - Std.$
Two	$\bar{x} - Std. \leq x < \bar{x} - 0.5Std.$
Three	$\bar{x} - 0.5Std. \leq x < \bar{x}$
.	.
.	.
M	$x \leq \bar{x} + Std.$

3. Formulate a frequency matrix labeled as $F = [f_{ij}]$, in which each element f_{ij} denotes the frequency of transitions from state i to state j within the random process during a single step.

$$F = \begin{bmatrix} f_{11} & f_{12} & .... & f_{1m} \\ f_{21} & f_{21} & ... & f_{2m} \\ ... & & ... & ... \\ f_{m1} & f_{m2} & ... & f_{mm} \end{bmatrix} \qquad (4)$$

4. Calculate the 1-step transition probability matrix, denoted as $P = [p_{ij}]$, where each element p_{ij} denotes the likelihood of moving from state i to state j in step (2). This matrix is statistically generated, taking into account different time lags, and is recorded as:

$$p^{(k)} = \begin{bmatrix} p_{11}^{(k)} & p_{12}^{(k)} & .... & p_{1m}^{(k)} \\ p_{21}^{(k)} & p_{22}^{(k)} & ..... & p_{2m}^{(k)} \\ .... & ..... & .... & \\ p_{m1}^{(k)} & p_{m2}^{(k)} & ... & p_{mm}^{(k)} \end{bmatrix} \text{ where } \quad k = 1,2,\dots,m$$

(5)

$$P_{ij} = \frac{f_{ij}}{\sum_{j=1}^{m} f_{ij}} \qquad (6)$$

5. Confirm the Markov property of the random process $\{X(t)\}$ by employing a Chi-Square test. To evaluate whether the five-state transition chain follows Markovian behavior or not, we apply the Chi-Square test using the following equation.[7],[9]

$$\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{m}\frac{(n_{ij}-n_{i.}n_{j.}/n)^2}{(n_{i.}n_{j.}/n)} \qquad (7)$$

If, when provided with the degrees of freedom (m - 1) × (m - 1) and α, the calculated χ^2 exceeds the critical value $\chi^2_{(\alpha,(m-1)\times(m-1))}$, then we reject the null hypothesis (H_0). In simpler terms, this implies that the random process shows a Markov property.

6. Calculate the upcoming output using a weighted Markov chain, where the weight of the Markov chain, represented as w_k, can be computed using the following formula:

$$w_k = \{w_1, w_2, \dots, w_m\}$$

$$w_k = \frac{ABS(r_k)}{\sum_{k=1}^{m} ABS(r_k)} \qquad (8)$$

Here, r_k signifies the ACF of the old data, with k ranging from 1 to m, and it can be computed using the subsequent equation:

$$r_k = \frac{\sum_{t=1}^{n-m}(x_t-\bar{x})(x_{t+k}-\bar{x})}{\sum_{t=1}^{n}(x_t-\bar{x})^2} \qquad (9)$$

where x_t denotes time series at time t, $\bar{x}$ denotes the average of time series x_t , and n denotes the no. of time series x_t. Thus, the formula of WMC is given by the following equation:

$$\hat{p}_{ij} = \sum_{k=1}^{m} w_k\,\hat{p}_{ij} \qquad , j=1,2,3,\dots,m \qquad (10)$$

$\hat{p}_{ij}$: Is the probability for a time series x_t to be in the state j in the future. The prediction outcome is in the form of a state, namely state j obtained by $\text{argmax}\{\hat{p}_{ij}, j = 1,2,\dots,m\}$.[5]

3.3 Time Series Modeling with the Box-Jenkins Approach

ARIMA Model is a commonly employed technique in time series forecasting and analysis. It combines three fundamental components: Autoregression (AR), Integration (I), and Moving Average (MA).

ARIMA is used to efficiently model and forecast time series data by capturing the inherent temporal connections, trends, and seasonal variations within the data. It finds utility across a range of fields such as economics, finance, epidemiology, and various other domains. To gain a deeper comprehension of ARIMA, let's explore the individual elements of this model:[1],[6]

1. **Autoregression (AR):** The autoregressive aspect pertains to how a present data point relates to its prior values. Within the ARIMA framework, the "AR" component signifies the reliance of the current data point on one or multiple preceding data points. The order of autoregression, denoted as "p," specifies how many past data points are considered when forecasting the current value.

$$y_t = \mu + \psi_1 y_{t-1} + \psi_2 y_{t-2} + \ldots\ldots + \psi_p y_{t-p} + \varepsilon_t \text{ where } i=1,2..., p \quad (11)$$

2. **Integration (I):** The integration element entails the process of differencing the time series to reduce it stationary. Stationarity means that the statistical characteristics of the series remain constant over time. Differencing is frequently employed to eliminate trends and seasonality, simplifying the modeling procedure. The degree of differencing, represented as "d," specifies how many iterations are necessary to transform the series into a stationary form.[3]
3. **Moving Average (MA):** The moving average aspect captures connection between a present data point and previous prediction errors, which are the disparities between forecasted and actual values. In ARIMA, the "MA" component models the short-term fluctuations around the overarching trend of the series. The order of the moving average, indicated as "q," defines how many prior

forecast errors are taken into account when anticipating the current data point.[3]

$$\hat{y}_t = \mu + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \ldots\ldots + \gamma_q \varepsilon_{t-q} + \varepsilon_t \text{ where } j=1, 2\ldots q \quad (12)$$

3.4 The notation for Box-Jenkins approach:

ARIMA (p, d, q), where p, d, q order of AR, I and MA respectively. Regarding general forecasting equation for the variable $\hat{y}_t$,can be expressed as:[11]

$$\hat{y}_t = \mu + \psi_1 y_{t-1} + \psi_2 y_{t-2} + \ldots\ldots + \psi_p y_{t-p} + \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \ldots\ldots + \gamma_q \varepsilon_{t-q} \quad (13)$$

The prediction equation is built as follows. Let y_t represent the d^{th} difference of Y, which means:

If $d = 0$, no stationary data then $y_t = Y_t$

If $d = 1$ then $y_t = Y_t - Y_{t-1}$

If $d = 2$ then $y_t = (Y_t - Y_{t-1})(Y_{t-1} - Y_{t-2})$ (14)

3.5 Steps to create an ARIMA model:

1. Data Preparation:

- Begin by cleaning and preprocessing your time series data.
- Assess the presence of trends and seasonality, and consider applying differencing to make the data stationary.[1]

2. Parameter Selection:

- Utilize autocorrelation and partial autocorrelation plots to estimate the values of AR order and MA order.

- Determine the value of differencing order by assessing how many differencing steps are required to achieve stationarity.[1]

3. Model Fitting:

- Fit the ARIMA (p, d, q) model to the differenced and stationary data.
- Employ optimization techniques to estimate the model's parameters.

4. Model Diagnostics:

- Examine the residuals of the model to identify autocorrelation and other patterns.
- If necessary, make adjustments to the model based on diagnostic results.

5. Forecasting:

- Utilize the established model to generate predictions for upcoming time intervals.

$$\hat{y}_t = \mu + \psi_1 y_{t-1} + \psi_2 y_{t-2} + \ldots\ldots + \psi_p y_{t-p} + \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \ldots\ldots + \gamma_q \varepsilon_{t-q} \qquad (15)$$

ARIMA models are robust tools for time series examination and prediction, yet they may not be well-suited for all data scenarios. Occasionally, more sophisticated approaches like Seasonal ARIMA (SARIMA), Exponential Smoothing techniques, or machine learning models could be better suited and more effective. [1]

3.6 Best fit model selection [1],[6]

Mean Absolute Error, Mean Absolute Percentage Error, and Root Mean Square Error are commonly used measures for evaluating accuracy of predictive models. These metrics assist in quantifying how closely a model's predictions align with the actual observed values. Here is a brief overview of each metric:

1. Mean Absolute Error (MAE): Calculates the mean of the absolute differences between forecast and real values. It measures the magnitude of errors without considering their way. A smaller MAE indicates greater accuracy.

Formula:

$$\mathrm{MAE}=\frac{1}{\mathrm{n}}\sum_{\mathrm{t}=1}^{\mathrm{n}}\mathrm{ABS}(\mathrm{e}_{\mathrm{t}}) \qquad (16)$$

2. Mean Absolute Percentage Error (MAPE): MAPE calculates the average percentage difference between predicted and actual values. It evaluates the relative accuracy of predictions in terms of percentage errors. MAPE is valuable when you want to understand the proportional magnitude of errors compared to the actual values

Formula:

$$\mathrm{MAPE}=\frac{1}{n}\sum_{t=1}^{n}\mathrm{ABS}\left(\frac{e_t}{y_t}\right) \times 100 \qquad (17)$$

3. Root Mean Square Error (RMSE): RMSE calculates the square root of the mean of the squared differences between predicted and actual values. It penalizes larger errors more significantly than smaller ones.

Formula:

$$\mathrm{RMSE}=\mathrm{SQRT}\left(\frac{1}{n}\sum_{t=1}^{n} e_t^2\right) \qquad (18)$$

In this context, where "y_t" represents the observed value at time "t," and "e_t" represents the discrepancy between the observed and predicted values, with "n" denoting the number of time points, smaller values for MAE, MAPE, and RMSE suggest a model that offers a better fit to the data.

4. The Akaike Information Criterion (AIC) is a statistical metric used in the domains of statistics and machine learning, primarily for the purpose of model selection. It was formulated in the 1970s by the Japanese statistician Hirotugu Akaike.

AIC considers two crucial components:[2],[6]

1. The model's likelihood, which measures how effectively the model fits the data.
2. The number of parameters within the model, serving as an indicator of the model's complexity.

The AIC is calculated using the following formula:

$$\text{Akaike Information Criterion } = n\ln(\hat{\sigma}^2) + 2M \tag{19}$$

In the equation 19:

- M: represents the number of estimated parameters within the model.
- $\hat{\sigma}^2 = (SSE/n)$

3.7 Application Parts and Data Description

Forecasting crude oil prices is a complex undertaking that requires a combination of various techniques and data inputs to achieve accurate predictions. Two commonly used approaches for time series forecasting in this context are the Weighted Markov Chain (WMC) and Autoregressive Integrated Moving Average (ARIMA) models. Here's a concise overview of how to apply both of these methods for forecasting crude oil prices:

1. Data Gathering and Preparation:

We obtain the data for crude oil price monthly in the website as follows: **https://sa.investing.com/commodities/brent-oil-historical-data,**

Figure (1): Time plot of price crude oil

<u>First:</u> Analysis of Weighted Markov Chain (WMC)

1. The data is organized into five distinct states based on price ranges, which are as follows:

- For prices less than 40, it falls into the "Low State" category.
- For prices between 40 (inclusive) and 65 (exclusive), it falls into the "Middle Low State" category.
- For prices between 65 (inclusive) and 90 (exclusive), it falls into the "Middle State" category.
- For prices between 90 (inclusive) and 115 (exclusive), it falls into the "Middle High State" category.
- For prices greater than or equal to 115, it falls into the "High State" category.

Table (2): Displays the frequency matrix depicting state transitions among the five price states in the crude oil price sequence.

	A(<40)	B[40,65)	C[65,90)	D[90,115)	E(>=115)	Sum
A(<40)	58	5	0	0	0	63
B[40,65)	4	67	10	0	0	81
C[65,90)	0	8	58	5	0	71
D[90,115)	1	1	3	50	5	60
E(>=115)	0	0	0	5	5	10

Sum	63	81	71	60	10	285

2. Test for the existence of a Markov chain

To determine if the five-state transition chain adheres to the characteristics of a weighted Markov chain, we employ a chi-square test as defined in equation (7):

Table (3): Chi-Square testing output of crude oil price

	j = 1	j = 2	j = 3	j = 4	j = 5
i = 1	$\frac{\left(58-\frac{63\times63}{285}\right)^2}{\left(\frac{63\times63}{285}\right)}$ =139.4834	9.301501	15.69473684	14.81053	0.663158
i = 2	10.79886	84.01648	5.134607254	17.05263	2.842105
i = 3	15.69474	7.35057	91.87617397	6.619904	2.491228
i = 4	11.33855	15.11127	9.549481097	110.5482	3.980263
i = 5	2.210526	2.842105	2.49122807	3.980263	61.60088
$\sum_{i=1}^{m}$	179.5261	118.6219	124.7462272	151.4642	73.125
$\sum_{i=1}^{m}\sum_{i=1}^{n}$	647.4834				

With "m" equal to 5, the degrees of freedom are calculated as (5-1)(5-1) = 16. Employing a significance level of 5% and consulting the chi-square table with 16 degrees of freedom, we identify $\chi^2_{(0.05,16)}$= = 26.296. The observed value of the sample statistic, chi-square, is 647.4834, which exceeds the chi-square table value. Consequently, we reject the null hypothesis that states are independent. This confirms that the state transition chain for crude oil prices from 2000 to 2023 follows a Markov chain.

3. Calculating Pearson autocorrelation coefficient:

We calculate the Pearson autocorrelation coefficients (r_k) using equation (9) and then proceed to calculate the Markov chain weights (w_k) using equation (8). In this examination, we explore prediction horizons ranging from one to five steps (L=1,2,3,4,5) for forecasting future crude oil prices. The output values for (r_k) and (w_k) are provided in Table 4 and 5, respectively.

Table (4): Estimated Autocorrelation Coefficients and Weight

k	k = 1	k = 2	k = 3	k = 4	k = 5
r_k	0.974963	0.936078	0.893542	0.852991	0.815267
$R_{k(Cumulative)}$	0.974963	0.974963+0.936078=1.911041	2.804584	3.657575	4.472842
w_k	(0.974963)/ 4.472842=0.217974	0.20928	0.199771	0.190705	0.18227

Table (5) The Weight Markov Chain of each (k) for every L=1,2,3,4,5

k	w_k				
	L = 1	L = 2	L = 3	L = 4	L = 5
1	0.974963/0.974963=1	0.936078/0.974963=0.510174	0.347632	0.26656	0.217974
2		0.936078/1.911041=0.489826	0.333767	0.255929	0.20928
3			0.318601	0.244299	0.199771
4				0.233212	0.190705
5					0.18227

Prior to forecasting the crude oil price for a particular month, our initial step involves predicting the oil price specifically for 01/10/2023 using equation (10). This is a vital stage for verifying the precision of the weighted Markov Chain. Equation (10) is formulated taking into account both the number of states, referred to as "m," and the maximum step, denoted as "L."

4. Cases of m=5, L=1,2,3,4,5

Table (6): The predicted price of crude oil in with m=1 and L=1,2,3,4,5

Months	i	k	w_k	W_kP_{ij}				
				j = 1	j = 2	j = 3	j = 4	j = 5
2023/05/01	3	1	0.217974	0	0.02456	0.178063	0.01535	0
2023/04/01	3	2	0.20928	0.001384	0.038988	0.14323	0.025238	0.00044
2023/03/01	3	3	0.199771	0.003414	0.046654	0.117361	0.031482	0.000859
2023/02/01	3	4	0.190705	0.005648	0.05017	0.098366	0.03535	0.00117

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (20) عدد (2) 2024

2023/01/01	3	5	0.18227	0.007842	0.051189	0.084234	0.037623	0.001381
				0.018289	0.211562	0.621254	0.145044	0.003851

Table (7): The predicted price of crude oil in with m=2 and L=1,2,3,4,5

Months	i	k	w_k	W_kP_{ij}				
				j = 1	j = 2	j = 3	j = 4	j = 5
2023/06/01	3	1	0.217974	0	0.02456046	0.17806331	0.01535	0
2023/05/01	3	2	0.20928	0.00138446	0.03898841	0.14322976	0.025238	0.000439942
2023/04/01	3	3	0.199771	0.0034141	0.04665391	0.11736124	0.031482	0.000859117
2023/03/01	3	4	0.190705	0.00564839	0.05016986	0.09836568	0.03535	0.001170496
2023/02/01	3	5	0.18227	0.00784236	0.05118921	0.08423397	0.037623	0.001381469
				0.01828931	0.21156185	0.62125396	0.145044	0.003851024

Table (8): The predicted price of crude oil in with m=5 and L=1,2,3,4,5

Months	i	k	w_k	W_kP_{ij}				
				j = 1	j = 2	j = 3	j = 4	j = 5
2023/09/01	4	1	0.217974	0.003253	0.003253	0.00976	0.195201	0.006507
2023/08/01	3	2	0.20928	0.001384	0.038988	0.14323	0.025238	0.00044
2023/07/01	3	3	0.199771	0.003414	0.046654	0.117361	0.031482	0.000859
2023/06/01	3	4	0.190705	0.005648	0.05017	0.098366	0.03535	0.00117
2023/05/01	3	5	0.18227	0.007842	0.051189	0.084234	0.037623	0.001381
				0.021543	0.190255	0.452951	0.324894	0.010358

Table (8) indicates that on 01/10/2023, the crude oil price is situated in state 3, specifically within the [65,90) range, with a probability of 0.452951. In this context, assuming m=5, it can be deduced that irrespective of the chosen value for L (which can range from 1 to 5), the crude oil price on 01/10/2023 will consistently fall within state 3, within the [65,90) interval.

5. Statistical Measure in WM-CM

The table presented below displays a collection of significant statistical metrics for the weighted Markov Chain model in comparison with ARIMA technique.

Table (9) Calculate statistical measures in WM-CM

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (20) عدد (2) 2024

Model	RMSE	MAE	MAPE	AIC
Weighted Markov chain model (WMCM)	10.3276	9.0725	10.2477	58.1120

<u>Second:</u> Analysis of ARIMA Model:

1. Prediction results for the crude oil price

Table (10) The different models for predicting the crude oil price

Groups	Type of models
(A)	Random walk
(B)	Constant mean= 66.5831
(C)	Linear trend= 45.1119 + 0.150148 t
(D)	Simple exp. smoothing with α= 0.9999
(E)	Brown's linear exp. smoothing with α = 0.6894
(F)	Holt's linear exp. smoothing with α = 0.9999 and β = 0.0084
(G)	ARIMA(2,0,0) with constant
(H)	ARIMA(1,1,0)
(I)	ARIMA(1,0,2) with constant
(J)	ARIMA(2,0,1) with constant
(K)	ARIMA(0,1,1)

The outcomes in Table (10), we observe that the time series model ARIMA (2,0,0) was selected between (11) models. The findings in Table (11) reveal that the optimal model for forecasting crude oil prices is (G), which demonstrates the highest performance standards in terms of predictive accuracy. The selected model is ARIMA (2,0,0).

$$y_t = \mu + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + e_t$$

$$y_t = 2.05896 + 1.18439 y_{t-1} - 0.213307 y_{t-2} + e_t$$

Table (11): Measures of Model Estimated

Model	RMSE	MAE	MAPE	AIC
(A)	6.27662	4.60183	7.6922	3.66458
(B)	29.171	24.1399	50.4323	6.74635
(C)	26.4728	22.1027	41.3241	6.55926
(D)	6.27032	4.61313	7.69533	3.67165
(E)	6.79545	5.06327	8.69847	3.83251
(F)	6.30572	4.59356	7.70045	3.68993
(G)	6.03433	4.46551	7.56419	3.61598
(H)	6.07962	4.53076	7.61319	3.6169

(I)	6.03225	4.4573	7.55473	3.62231
(J)	6.03283	4.46898	7.59002	3.6225
(K)	6.09679	4.53849	7.61311	3.62254

2. Table (12) provides an overview of the elements comprising the ARIMA (2,0,0) model, which was selected based on the forecasting criteria.

Table (12): ARIMA Model Summary

Parameter	Estimate	Stnd. Error	t	P-value
AR (1)	1.18439	0.0575052	20.5962	0.000
AR (2)	-0.213307	0.0574532	-3.71271	0.000
Mean	71.1975	11.9371	5.96438	0.000
Constant	2.05896	0.345	5.968	0.000

Table (12) reveals that the parameters of the selected ARIMA model, ARIMA (2, 0, 0), exhibit statistical significance. This signifies the model's validity for predicting future time periods. Additionally, Figure (3) illustrates the comparison between actual and predicted crude oil price values according to the chosen model.

3. Three tests were conducted to assess whether the residuals constitute a random sequence. Such a sequence of random numbers is commonly referred to as "white noise" because it exhibits uniform contributions across multiple frequencies.

(1) Runs above and below median

Median = 0.20899

Number of runs above and below median = 142

Expected number of runs = 143.0

Large sample test statistic z = 0.0594442

P-value = 0.952593

(2) Runs up and down

Number of runs up and down = 195

Expected number of runs = 189.667

Large sample test statistic z = 0.681194

P-value = 0.495746

(3) Box-Pierce Test

Test based on first 24 autocorrelations

Large sample test statistic = 12.3115

P-value = 0.950608

The initial test involves tracking how many times the sequence crosses above or below the median. In this instance, there are 142 such occurrences, while the expected value for a random sequence is 143.0. Given that the P-value for this test is equal to or greater than 0.05, we cannot dismiss the hypothesis that the residuals show randomness with a confidence level of 95.0% or higher.

The second examination counts the occurrences of the sequence rising or falling. There are 195 such instances, in contrast to an expected value of 189.667 for a random sequence. Since the P-value for this test is greater than or equal to 0.05, we are unable to reject the idea that the series displays randomness at a confidence level of 95.0% or higher.

The third assessment centers on the sum of squares of the initial 24 autocorrelation coefficients. Because the P-value for this test is greater than or equal to 0.05, we cannot reject the hypothesis that the series demonstrates randomness at a confidence level of 95.0% or higher.

4. Test for Normality of Residual

Hypothesis Test:

H_0: Residual is Normally distributed vs. H_a : Residual is Non-Normality distribution

مجلة الغري للعلوم الاقتصادية والادارية
مجلد (20) عدد (2) 2024

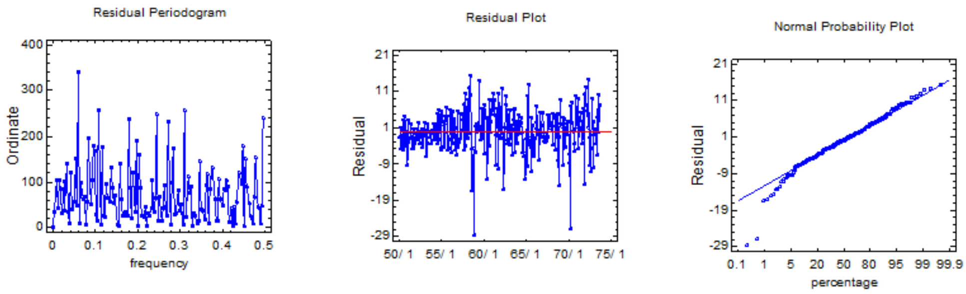

Figure (2): Display the residuals to conduct a test for normality

Upon examining Figure No. (2), it is clear that the residuals derived from the data conform to a normal distribution. As a result, we validate the null hypothesis, which asserts that the residuals indeed exhibit a normal distribution

.5. Comparison Between weighted Markov chain (WMC) and ARIMA Model

Table (9) indicates that the most accurate forecasting performance is achieved with the ARIMA (2,0,0) model, exhibiting the lowest MAPE of 7.56419 and RMSE of 6.03433, demonstrating an excellent fit in comparison to the Weighted Markov Chain (WMC).

Table (13) Shows the different two models

Model	RMSE	MAE	MAPE	AIC
Weighted Markov chain (WMC)	10.3276	9.0725	10.2477	58.1120
ARIMA(2,0,0)	6.03433	4.46551	7.56419	3.61598

6.Generating Forecasting

After completing the diagnostic phase and estimating parameters for our proposed model, we transition to the prediction stage. This phase marks the final step in the analysis and selection of an appropriate time series model and is a key objective when building statistical models for time series data. In this context, we utilized the proposed model to forecast crude oil prices and evaluated the model's reliability by considering 95% confidence intervals. In the realm of statistical modeling, it's crucial to assess both the precision of our predictive values and the confidence intervals that accompany these forecasts.

Table (14): Forecast Table for Price of Crude Oil

Period	Forecast	Lower 95.0% Limit	Upper 95.0% Limit
10	94.7871	82.9024	106.672
11	94.2877	75.8653	112.71
12	93.5135	70.2922	116.735
1	92.7029	65.7209	119.685
2	91.9081	61.8445	121.972
3	91.1397	58.4753	123.804
4	90.399	55.4958	125.302
5	89.6858	52.8281	126.543
6	88.999	50.4176	127.58
7	88.3377	48.2245	128.451
8	87.7009	46.2181	129.184
9	87.0879	44.3745	129.801

The table above summarizes the projected crude oil price values, incorporating forecasts generated by the model and the discrepancies (the difference between actual data and forecasts) observed during the available data timeframe. For time periods extending beyond the dataset, it provides prediction limits at a 95.0% confidence level for the forecasts. These limits offer a sense of the probable range of actual data values in future periods, assuming that the model used is appropriate for the data. To visualize the forecasts, you can opt for the "Forecast Plot" feature in the graphical options and modify the confidence level as you view the plot by right-clicking and selecting "Pane Options." Additionally, to assess the model's accuracy, you can choose "Model Comparisons" from the tabular options.

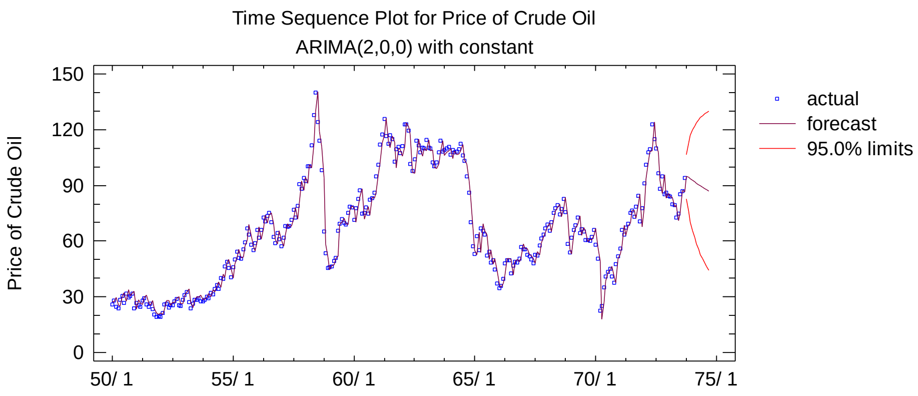

Figure (3): Time Series Plot for Price of Crude Oil ARIMA (2,0,0) with constant

4 Conclusions and Recommendations

Predicting crude oil prices is of utmost significance for a diverse set of stakeholders, encompassing investors, policymakers, and industry experts. To enhance the precision of these predictions, this study investigated the utilization of two distinct methodologies: the Weighted Markov Chain (WMC) and the Autoregressive Integrated Moving Average (ARIMA) models. Drawing from the outcomes and analysis of these approaches, we can draw the following conclusions and recommendations:

4.1 Conclusions

1. Based on the analysis of the results, the autoregressive integrated moving average (ARIMA) model with parameters (2, 0, 0) proved to be a successful method for predicting crude oil prices.

2. The model of the analysis yields an ARIMA model with parameters (2,0,0).

$$y_t = 2.05896 + 1.18439y_{t-1} - 0.213307y_{t-2} + e_t$$

3. The forecasted results indicate an uptrend in crude oil prices for the months 10 and 11, followed by a downtrend for months 9 through 12 Consequently, it is imperative to implement price adjustments and maintain vigilant monitoring of oil prices.
4. Within the weighted Markov chain model (WMCM), the probability of crude oil prices falls within the third tier, which ranges from 65 to 90. This outcome closely aligns with the findings derived from the ARIMA (2,0,0) model.

4.2 Recommendations

Based on the conclusions of the study, it is recommended the followings:

1. Integrate the WMC model with complementary forecasting methods like ARIMA to boost its predictive capabilities and address its shortcomings.
2. Consider incorporating exogenous variables, such as geopolitical events, economic indicators, and energy production data, to enhance the model's predictive accuracy.
3. Investing in crude oil and related assets carries inherent risks, and past performance is not indicative of future results. It's crucial to conduct thorough research, diversify your investments, and carefully consider your risk tolerance and investment goals when dealing with crude oil prices.

References

[1] Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). **Time Series Analysis: Forecasting and Control.** This classic book by George Box and others provides an in-depth introduction to time series analysis and includes extensive discussions of ARIMA models.

[2] Enders, W. (2014)**. Applied Econometric Time Series.** William Enders' book covers various time series models, including ARIMA, and provides practical examples and applications.

[3] Gao Gao, Kwoklun Lo, Fulin Fan (2017)**. Comparison of ARIMA and ANN Models Used in Electricity Price Forecasting for Power Market,** Energy and Power Engineering Vol.9 No.4B

[4] Hyndman, R. J., & Athanasopoulos, G. (2018)**. Forecasting: Principles and Practice.** This online textbook by Rob J. Hyndman and George Athanasopoulos includes a comprehensive section on ARIMA modeling, complete with examples and exercises.

[5] Hamdin, A. O., & Hussein, M. M. (2023)**. An application of Wavelet Markov Chains Model to Study Earthquake Occurrence**. Tikrit Journal of Administrative and Economic Sciences, 19(61, 2), 334-353.

[6] Hussein, M. M. F. (2022)**. Comparison of Time Series Models before and after Using Wavelet Shrinkage Filtering to Forecast the Amount of Natural Gas in Iraq**. Cihan University-Erbil Scientific Journal, 6(1), 32-46.

[7] Hussein, M. M. F. (2023). **Predicting Number of People Living with Chronic HCV Using Gray-Weighted Markov Chain Model (GW-MCM**). University of Kirkuk Journal For Administrative and Economic Science, 13(4).

[8] J. Zhang, W.X. Tao, Q. Wang (2016)**. Application of Weighted Markov Chain in Forecasting Rainfall in Jinan City[J]**. Yellow River, 38(9): 13-16. (In Chinese)

[9] P. Zhi-Hang, X. Le-Tian and S. Yong-Mei (2005). **Application of Weighted Markov Chain in the Prediction of Year's Harvest of Crops**, Mathematics in Practice and Theory, pp.30-35.

[10] Y. B.,Li, , Xu, P. P., Qian, H., & Wang, H. (2017). **Predicting Precipitation in Tongchuan Using Weighting Markov Chain Model [J]**. Journal of Irrigation and Drainage, 5, 96-102.

[11] Zhou, Y. (2017)**. Application of ARIMA Model in Forecasting and Analyzing Oil Price**. Northern Economy and Trade, No. 8, 23-24.