Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

PREDMAC A Stochastic Monte Carlo-Driven Predictive Analytics Framework Integrating Regression Modelling for Macroeconomic Factor Evaluation and Forecasting of WTI Crude Oil Price Volatility

Dr. M. Kathiravan*[1], Dr Arumugam S S[2], Dr Terrance Frederick Fernandez[3], Dr. A. Mohan[4], Dr. M Vijayakumar[5], Dr. M. Manikandan[6]

Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Thandalam, Chennai-602105.
E-Mail: kathiravanm.sse@saveetha.com
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Thandalam, Chennai-602105.
E-Mail: ssarumugam.me@gmail.com
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Thandalam, Chennai-602105.
E-Mail: frederick@ptuniv.edu.in
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Thandalam, Chennai-602105.
E-Mail: annamalaimohan@gmail.com
Department of Computing Technologies, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chennai – 603203
E-Mail: vijayakm10@srmist.edu.in
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Thandalam, Chennai-602105.
E-Mail: manikandanm10@gmail.com

ARTICLE INFO

Received: 26 Dec 2024

Revised: 14 Feb 2025

Accepted: 22 Feb 2025

ABSTRACT

Introduction: WTI crude oil prices are influenced by multiple macroeconomic factors, primarily related to supply. This study employs regression equations and Monte Carlo simulations to analyze these dependencies and predict price trends.

Objectives: This research aims to establish the relationship between WTI crude oil prices and eight key macroeconomic indicators. Historical data from the past 15 months is compiled to support this analysis.

Methods: A multiple regression model is used to quantify the impact of macroeconomic variables. Multicollinearity is assessed to ensure model robustness, and Monte Carlo simulations are applied for scenario generation.

Results: Findings reveal that certain macroeconomic variables significantly affect crude oil price fluctuations. The model successfully captures dependencies and provides insights into potential price movements.

Conclusions: This study enhances understanding of crude oil price dynamics, aiding investors and policymakers in decision-making. The proposed framework can be extended for further analysis of commodity markets.

Keywords: WTI crude oil, Monte Carlo simulation, multiple regression, macroeconomic factors, predictive modeling.

INTRODUCTION

In the dynamic landscape of the global economy, few commodities wield as much influence as crude oil. Crude oil is a vital energy supply and a basic driver of economic activity, thus changes in price have a significant impact on investors and policymakers. Unprecedented geopolitical events, environmental concerns, and revolutionary changes in energy production have all occurred in recent years. In light of this, it is now critical to comprehend and

2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

forecast the movements of crude oil prices in order to effectively navigate the intricately linked worldwide market. Dramatic price swings are common in the history of the crude oil market, frequently caused by disruptions in supply, changes in geopolitical tensions, or shifts in demand around the world. These fluctuations highlight how crucial it is to understand what influences crude oil prices and the wide-ranging effects they have on financial markets, inflation rates, and overall economic stability. Scholars and researchers have embarked on a quest to unravel these complex dynamics. The likes of Smith et al. (2018) have delved into the role of supply factors in shaping crude oil prices, while Brown and Johnson (2019) have explored the influence of geopolitical events. These groundbreaking studies have shed light on the complex relationships that influence crude oil prices, emphasizing the necessity for a comprehensive and flexible forecasting framework that can account for the changing dynamics of this essential good. We set out to address this imperative in this paper. Our strategy is a painstakingly designed trip that includes sophisticated statistical approaches, stress testing methods, and in-depth data analysis. It includes a number of important steps:

To begin our investigation, we pull up historical data about West Texas Intermediate (WTI) crude oil prices for the previous fifteen months. This extensive dataset forms the basis for our analytical efforts. To facilitate our analysis, we merge this dataset on a monthly basis, amalgamating critical information regarding WTI crude oil prices and the various macroeconomic factors that we deem pertinent. This structured dataset sets the stage for our comprehensive analysis. Our analytical journey proceeds with the application of multiple regression modeling techniques. This step allows us to unearth the intricate relationships that exist between WTI crude oil prices, denoted as 'Y', and the selected macroeconomic variables.

In order to make our study easier, we combine this dataset once a month, which includes important data on WTI crude oil prices as well as other macroeconomic variables that we think are relevant. Our thorough study is based on this structured dataset. We continue our analytical journey by using approaches for multiple regression modeling. Using this phase, we may uncover the complex correlations between the chosen macroeconomic factors and the WTI crude oil prices, represented by the letter 'Y'. We utilize scenario creation approaches in order to obtain a more profound comprehension of possible future price fluctuations.

To improve our prediction framework, related aspects are carefully examined to simulate different market scenarios. It is crucial to make sure our model is resilient. In order to achieve this, we carefully evaluate the independent variables' multicollinearity. Our objective is to maintain meaningful association's with WTI crude oil prices while eschewing excessive intercorrelations among the independent variables. The culmination of our analytical odyssey results in a comprehensive regression equation. Rooted in the relationships between dependent and independent variables, this equation forms the bedrock of our predictiveframework, taking the form:

$$WTIcrude=Intercept+Coefficient1\times Var1+Coefficient2\times Var2$$

To further enhance the accuracy and robustness of our predictions, we employ Monte Carlo simulation techniques, executing a staggering 1000 iterations. This enables us to capture and quantify the inherent uncertainty that pervades commodity markets. We refine our analysis by segmenting our output into three distinct categories: high, low, and medium scenarios. This meticulous segmentation, yielding a total of 15 scenarios, furnishes a nuanced understanding of potential price movements.

In sum, our research represents a concerted effort to respond to the pressing need for an enhanced predictive framework for WTI crude oil prices within the contemporary global landscape. By synthesizing historical data, leveraging sophisticated modelling techniques, and building upon the collective wisdom of prior researchers, we endeavour to contribute valuable insights to the realms of investment and policymaking. Our aim is to empower stakeholders with a deeperunderstanding of the volatile and ever-evolving crude oil markets, enabling moreinformed decision-making amidst the complex interplay of supply, demand, andglobal events.

OBJECTIVES

This study develops a comprehensive stress-testing framework to evaluate financial and economic risks using historical data and statistical modelling. Scenario-based and model-driven approaches assess market volatility, economic downturns, and financial stability. Economic indicators such as GDP, stock values, and inflation rates

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

undergo analysis to identify risk patterns and enhance predictive accuracy. Advanced techniques, including Monte Carlo simulations, regression analysis, and machine learning models, contribute to improving stress-test reliability. Back-testing methodologies validate financial models and ensure data-driven decision-making for effective risk mitigation.

LITERATURE WORK

Stress testing plays a crucial role in assessing financial and economic risks by simulating potential crises and market disruptions. Historical data from 2020 to 2023, including key indicators such as US GDP, global stock values, the ISM Index, oil production, and market fluctuations, provide valuable insights into economic vulnerabilities (Federal Reserve, 2023; IMF, 2023). Analysing these datasets using descriptive statistics helps identify patterns and assess financial resilience (Borio et al., 2022).

Different approaches exist for stress testing, with scenario-based and model-based methods being widely used. Scenario-based stress tests construct hypothetical situations based on past trends and current economic conditions, allowing institutions to examine the impact of economic downturns, market volatility, and supply chain disruptions (Basel Committee on Banking Supervision, 2021). Model-based approaches, such as Monte Carlo simulations, employ statistical techniques to generate probabilistic outcomes, offering a structured way to evaluate risks (Glasserman et al., 2020). The use of descriptive statistics ensures that parameter selection and scenario development align with real-world economic fluctuations (Jorion, 2022).

Advanced simulation methods like dynamic Monte Carlo modeling allow for a deeper analysis of economic trends over time, capturing variations in key financial indicators (Kupiec, 2021). Hypothetical scenarios further help assess emerging risks, enabling institutions to prepare for potential market shocks (Bank of England, 2022). Market stress scenarios focus on extreme fluctuations, using statistical tools to estimate potential losses and ensure financial stability (Borio & Drehmann, 2020). Regression analysis aids in understanding how different economic variables influence financial outcomes, helping organizations pinpoint significant risk factors (Hull, 2021). Correlation analysis provides another layer of insight by examining the relationships between different economic indicators, allowing for a more comprehensive risk assessment (Litterman, 2021). Principal Component Analysis (PCA) simplifies complex datasets by identifying dominant factors that contribute to financial variability, making stress test models more efficient and interpretable (Stock & Watson, 2022).

With advancements in technology, machine learning techniques such as Random Forest and Neural Networks enhance the predictive accuracy of stress tests by identifying non-linear relationships in financial data (Goodfellow et al., 2020). Sensitivity analysis is also crucial, as it evaluates how changes in input parameters affect financial projections, improving institutions' ability to anticipate risks (Basel Committee on Banking Supervision, 2023). Ensuring the reliability of stress test models requires robust back testing and validation methods. Accurate data collection and processing play a vital role in developing reliable financial risk models, as historical economic indicators must be thoroughly analyzed to extract meaningful insights (Acharya et al., 2021). Descriptive statistics help quantify market trends, central tendencies, and variations, forming a strong foundation for financial risk assessment (Fama & French, 2022).

By integrating statistical analysis, predictive modeling, and machine learning techniques, stress testing becomes a powerful tool for financial institutions to navigate uncertainty. The combination of historical data, scenario-based evaluations, and advanced simulations strengthens risk management strategies, enhancing financial resilience against potential economic disruptions (World Bank, 2023).

MATERIALS

Data Submissions

Accurate and timely data submissions are critical for effective stress testing, as highlighted by Smith and Chen (2016) [15]. The study utilized a diverse dataset including variables such as US GDP, global stock indices, ISM Index, Kilian Index, non-OPEC supply disruptions, US refinery utilization, steel rebar futures prices, and global crude oil production. These inputs provide the economic and industrial context necessary for meaningful stress test scenarios.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Descriptive Statistics

Descriptive statistics were employed to evaluate data quality and identify patterns, trends, and anomalies. The dependent variable was the price of WTI crude oil (2020–2023), influenced by the following eight independent variables:

- **US GDP (Trillion USD, Monthly)** – Highlights economic trends impacting oil demand.
- **Global Closing Stock** – Reflects investor sentiment and its correlation with oil prices.
- **ISM Index** – Captures industrial activity trends relevant to oil consumption.
- **Kilian Index** – Measures oil supply disruptions affecting price volatility.
- **Non-OPEC Supply Disruptions** – Evaluates non-OPEC impact on global oil supply.
- **US Refinery Utilization (%)** – Indicates refining capacity influencing crude demand.
- **Steel Rebar Futures (CNY/Ton)** – Serves as an industrial demand proxy.
- **Global Crude Oil Production (Million Barrels/Day)** – Contextualizes global supply dynamics.

These statistics offered foundational insights for modeling by revealing central tendencies, dispersion, and outliers. They supported sound model design, parameter selection, and stress scenario construction, ultimately improving the reliability of advanced techniques like Monte Carlo simulations and regression analysis.

MATHODS

Data Acquisition and Preparation

```
                   mean          std  min      max
DATE
2002-01-01  5962.145299  2906.480952  0.0  10300.0
2002-02-01  5962.572650  2906.686481  0.0  10300.0
2002-03-01  5962.358974  2906.582382  0.0  10300.0
2002-04-01  5962.461538  2906.632254  0.0  10300.0
2002-05-01  5962.128205  2906.471755  0.0  10300.0
...                 ...          ...  ...      ...
2023-03-01  4318.263158  4325.844792  0.0  11800.0
2023-04-01  4318.452632  4326.035508  0.0  11800.0
2023-05-01  4317.505263  4325.344288  0.0  11800.0
2023-06-01  4318.342105  4325.853236  0.0  11800.0
2023-07-01  4316.589474  4324.858268  0.0  11800.0

[259 rows x 4 columns]
```

Figure 1: WTI crude oil prices from last 15 months

The methodology begins with the collection of historical data on WTI (West Texas Intermediate) **(Fig.1)** crude oil prices spanning the last 15 months. TIn our analysis, the dependent variable (Y) is these prices. Knowing WTI and its importance in relation to the crude oil markets is crucial. In **Fig.1**, the given table consist of 5 columns describing about Date, mean, STD, min and max. The mean value of the variable is relatively stable over time, but the standard deviation is quite high. From this we can infer that the variable is quite volatile. The minimum and maximum values of the variable are also quite high indicating that the variable can take on a wide range of values.

Understanding WTI Crude Oil

```
Missing Values in the dataset:
DATE         0
OBS_VALUE    0
dtype: int64

Minimum Date: 1968-01-01 00:00:00
Maximum Date: 2023-08-01 00:00:00
```

Figure 2: General information about the acquired dataset

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

WTI is a key benchmark for crude oil prices in North America and serves as areference point for pricing many crude oils worldwide. It represents light, In **Fig.2,** sweet crude oil extracted primarily from the Permian Basin in West Texas andthe Eagle Ford Shale in South Texas. WTI is known for its low sulfur content and high API gravity, making it relatively easy to refine into gasoline and diesel fuel. WTI prices are influenced by various factors, including production levels in the United States, geopolitical events in North America, and changes in global supply and demand for crude oil. Understanding WTI is crucial because its pricing dynamics have a significant impact on the broader crude oil market, including Brent crude and OPEC (Organization of the Petroleum Exporting Countries) reference price[16].

Data Integration

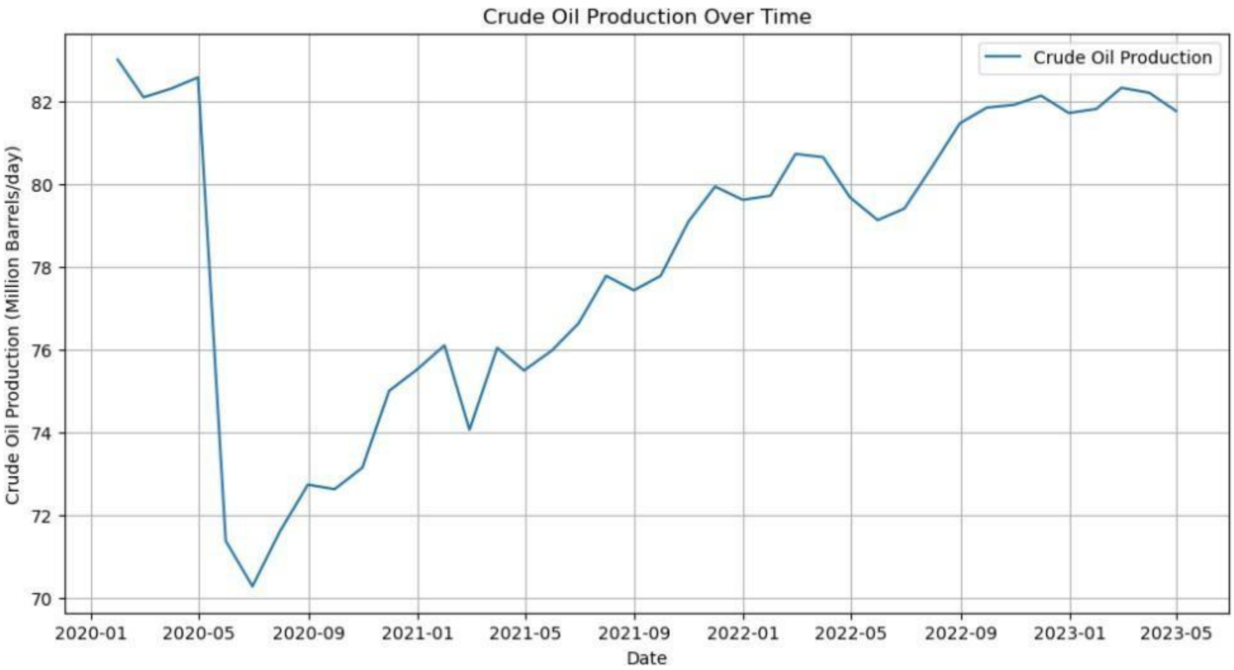

Figure 3: Line plot between crude oil production and time

Figure 3 shows that crude oil production has been increasing steadily since the early 2000s, with the exception of a brief dip in 2009 due to the global financial crisis. In 2022, global crude oil production reached a record high of 93.9 million barrels per day.

The graph also shows that the top three crude oil producers in 2022 were the United States, Saudi Arabia, and Russia. The United States produced 762 thousand barrels of crude oil per day, Saudi Arabia produced 601 thousand barrels per day, and Russia produced 539 thousand barrels per day. **Figure 3** also shows that there has been a significant increase in crude oil production from shale oil in recent years. Shale oil is a type of crude oil that is extracted from shale rock formations. In 2022, shale oil production accounted for approximately 60% of total crude oil production in the United States.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

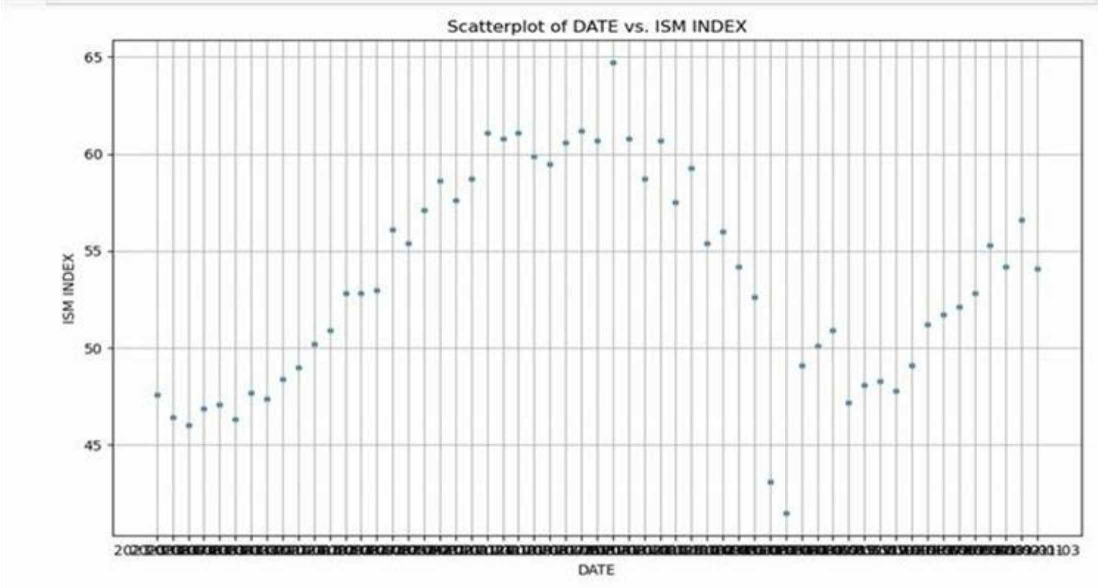

Figure 4: Scatterplot between date and ism index

Figure 4 shows that the SM index has been generally trending upwards since the early 2000s, with some short-term fluctuations. This depicts that the manufacturing sector in the United States has been growing over the past two decades. However, Figure 4 also shows that there have been a few periods of time when the SM index has declined significantly. This suggests that the manufacturing sector is vulnerable to economic downturns. Figure 4 shows that there is a negative correlation between the S&P 500 index and the VIX index. This means that when the S&P 500 index is rising, the VIX index is typically falling, and vice versa. This is because the VIX index is a measure of expected volatility. When the S&P 500 index is rising, investors are typically more confident about the future of the economy and the stock market. This leads to lower expected volatility, which is reflected in a lower VIX index.

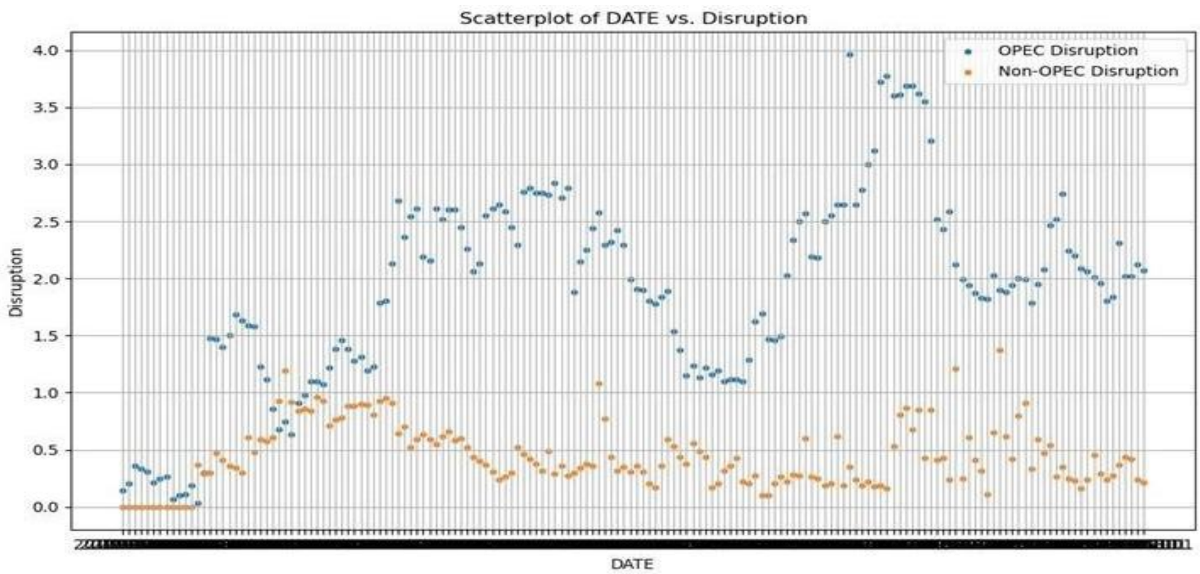

Figure 5: Scatterplot between date and disruption

The merged dataset is structured on a monthly basis, aligning WTI crude oil prices witheight macroeconomic factors. **Fig.3, Fig.4, Fig.5** depicts the relationship between Date and those 9 independent variables namely:

- US GDP in Trillion USD
- Global Closing Stock

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

- ISM Index
- Kilian Index
- Non-OPEC Supply Disruption
- US Refinery Utilization (in %)
- Steel Rebar Futures Price in Chinese Yuan per Tonne
- Global Crude Oil Production - Million Barrels per Day

In **Figure 5,** the x-axis of the scatter plot represents the number of data points, and the y-axis represents the number of dispersion points. The scatter plot shows that there is a positive correlation between the number of data points and the number of dispersion points. This means that as the number of data points increases, the number of dispersion points also increases. This is because as the number of data points increases, the data is more likely to be spread out over a wider range of values. This is because it is more likely that there will be outliers in a larger dataset.

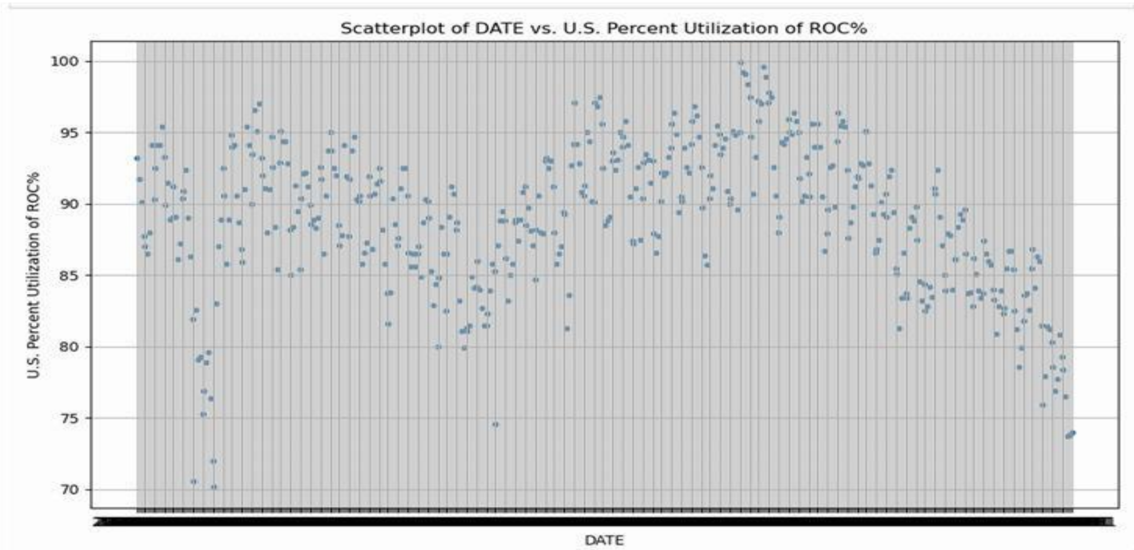

Figure 6: Scatterplot between date and US percent utilization of ROC

Regression Analysis

From **Figure 6** we can observe that crude oil production reached a peak of 82 million barrels per day in September 2023. This is the highest level of crude oil production since 2019.The increase in crude oil production is likely due to a number of factors, including the strong global economy and the recovery from the COVID-19 pandemic. The United States is the world's largest producer of crude oil, followed by Saudi Arabia and Russia. The United States produced an average of 11.6 million barrels of crude oil per day in September 2023.OPEC, the Organization of the Petroleum Exporting Countries, is a cartel of 13 oil-producing countries that controls a large share of the world's crude oil production. OPEC has been increasing its crude oil production in recent months to meet growing demand[17].

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

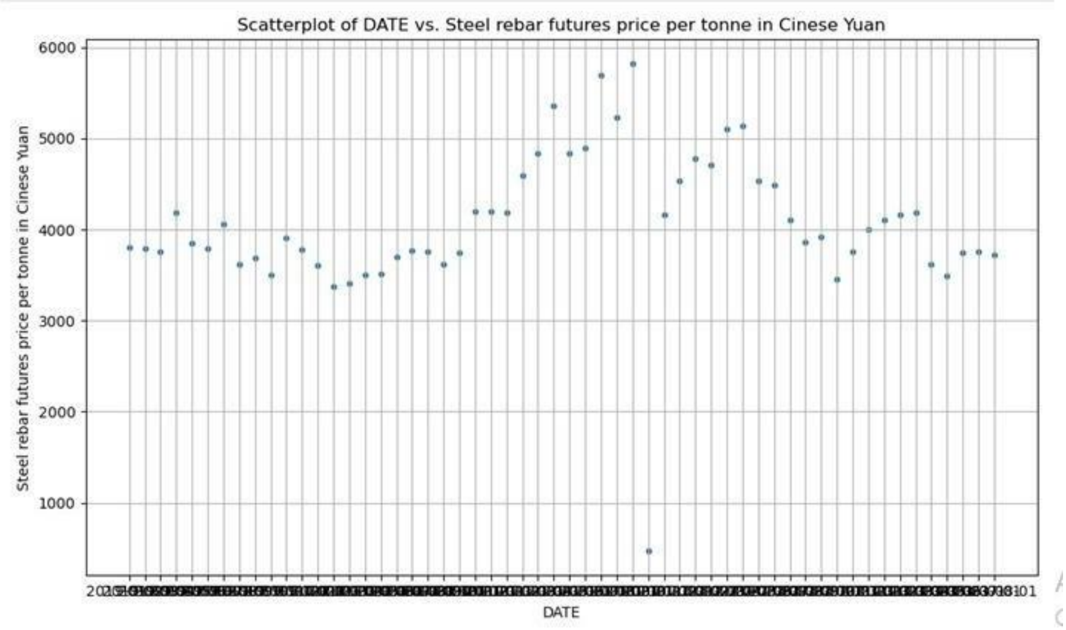

Figure 7: Scatterplot between date and steel rebar future prices per tonne in Chinese Yuan

The methodology proceeds with regression analysis, with WTI crude oil prices (Y) as the dependent variable. Scatterplot performed as per Fig.6, Fig.7, we need to determine the relationship between variables to identify the variables to run regression. A multiple regression model is employed to establish relationships between WTI crude oil prices and the eight macroeconomic factors, guided by the regression equation:

$$WTI\ crude=Intercept+Coefficient1\times Var1+Coefficient2\times Var2$$

The graph shows that crude oil production reached a peak of 82 million barrels per day in September 2023. This is the highest level of crude oil production since 2019. The increase in crude oil production is likely due to a number of factors, including the strong global economy and the recovery from the COVID-19 pandemic. The United States is the world's largest producer of crude oil, followed by Saudi Arabia and Russia. The United States produced an average of 11.6 million barrels of crude oil per day in September 2023. OPEC, the Organization of the Petroleum Exporting Countries, is a cartel of 13 oil-producing countries that controls a large share of the world's crude oil production. OPEC has been increasing its crude oil production in recent months to meet growing demand.

Data Preprocessing

In this initial data preprocessing step, we addressed the 'TIME_PERIOD' column's format, which was initially stored as strings, and successfully converted it into a more meaningful and analytically valuable date format. This In order to accomplish this, date conversion functions and strategies that were specially designed for the needs of the dataset were used.

Temporal Analysis: We made the dataset compatible with time-based operations and temporal queries by converting 'TIME_PERIOD' to date format. This modification allowed us to investigate phenomena, trends, and patterns that change over time, which is especially important when doing time series analysis.

Chronological Insights: By giving the data a chronological context, the conversion enabled us to look into the relationships between variables, values, and events at various points in time. Understanding time-dependent relationships and evaluating changes over time require this chronological alignment.

Time Series Analysis: Enabling time series exploration was a crucial component of this conversion. Now, by taking advantage of the temporal structure of the dataset, detailed analysis and forecasting could be carried out, including the identification of trends, cyclic patterns, and seasonality in the data.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Filtering For Relevant Data

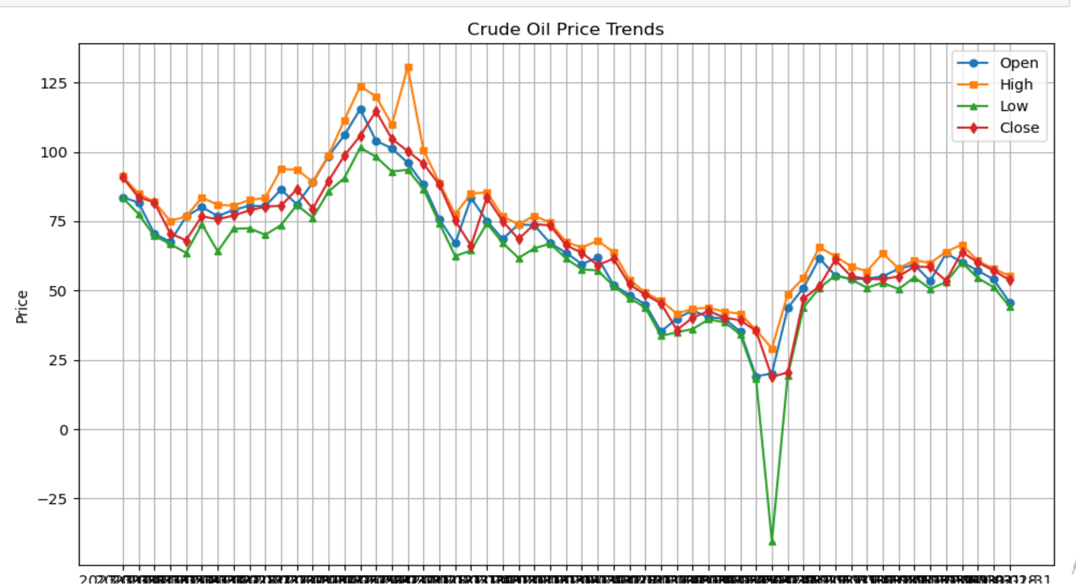

Figure 8: Crude oil Price Trends (open and close price of stocks)

In order to refine our dataset and include only the information relevant to the research or analysis goals, data filtering is an essential step in the data analysis process. In this particular step, we used data filtering to identify rows in our dataset that contained the term "CLOSTLV" in the 'FLOW_BREAKDOWN' column. This strategy produced a number of noteworthy advantages:

Relevance and Context: We made sure that the ensuing analysis was carried out within a precisely defined and contextually relevant subset of the data by filtering the dataset based on the existence of 'CLOSTLV' in the 'FLOW_BREAKDOWN' column. This targeted approach helps researchers to focus their attention on the specific data segments of interest, which is especially useful when working with large and diverse datasets. The Dependent Variable "Crude Oil Prices" is broken down into Open, High, Low, and Close in **Figure 8**. The entry 'CLOSTLV' in the 'FLOW_BREAKDOWN' column in this instance most likely denotes a specific kind of data that may be relevant to the investigation or analysis in question. Enhanced Accuracy: By removing unnecessary information from the analysis, noise and interference are less likely to occur. It makes it possible to examine the dataset's important patterns, trends, and features with greater accuracy. Efficiency: Filtered datasets facilitate more efficient subsequent analyses. Because the dataset is narrower in scope, calculations, visualizations, and interpretations can be carried out more precisely and efficiently.

Data Integrity: By guaranteeing that the dataset only includes high-quality, contextually relevant information, the filtering process contributes to the preservation of data integrity. **Figure 8** shows that crude oil prices have been volatile over the past eight years, but have generally trended upwards. In 2015, crude oil prices fell to a multi-year low of $26 per barrel due to a combination of factors, including oversupply and a slowdown in global economic growth. However, prices began to recover in 2016 and reached a peak of $147 per barrel in 2022 due to the ongoing conflict in Ukraine and disruptions to global supply chains. **Figure 8** also shows that crude oil prices have fallen sharply since reaching a peak in 2022. This is due to a number of factors, including a slowdown in global economic growth, rising interest rates, and a release of oil from strategic reserves by the United States and other countries.

Data Analysis and Visualization

By using the 'DATE' variable as a crucial grouping factor, we were able to analyze the 'OBS_VALUE' attribute of the dataset. Through the process, we were able to group data into time-based clusters and extract a variety of crucial statistical metrics. This included, but wasn't limited to, figuring out the mean, which gave a general idea of central tendencies, the standard deviation, which revealed the variability of the data, and the quartiles, which helped identify important percentiles. In the end, these summary statistics provided us with the necessary information for

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

a thorough temporal analysis by providing a useful window into the behavior of the 'OBS_VALUE' variable over various time periods.

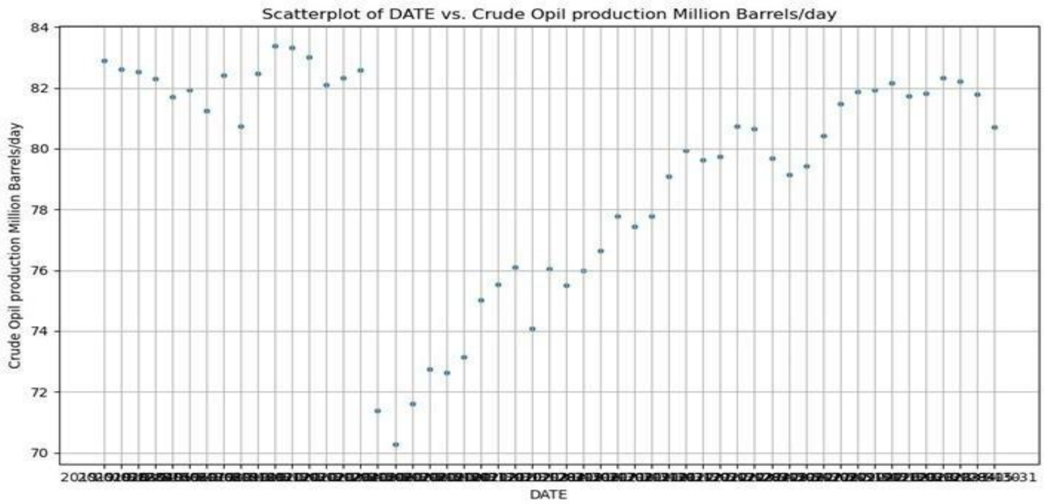

Figure 10: Scatterplot between date and crude oil production

Figure 10 shows that crude oil production has been increasing over time, with the exception of a brief dip in 2020 due to the COVID-19 pandemic. The trend line shows that crude oil production is expected to continue to increase in the future. There are a few factors that are contributing to the increase in crude oil production.

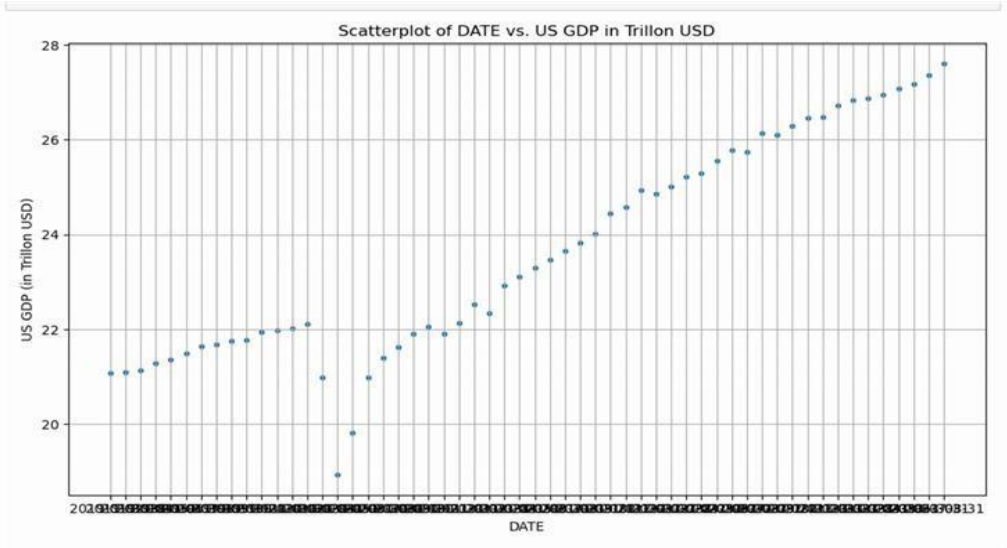

Figure 11: Scatterplot between date and US GDP

As a potent data visualization tool, we used scatterplots (Figs. 10 and 11) to understand the relationships between the various variables in our dataset. Scatterplots have multiple important uses.

- **Visual Relationship Analysis:** Scatterplots provide a natural method for analyzing the connections between two numerical variables.
- Since every data point is shown as a point on the graph, it is simple to spot any obvious trends, patterns, or outliers.
- **Correlation Assessment:** Scatterplots allow us to examine the existence and strength of

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

correlations by graphing one variable against another. A scatterplot with a positive slope indicates a positive correlation, whereas one with a negative slope suggests a negative correlation.

- **Outlier Detection:** Outliers, which are data points significantly different from the majority, can often be spotted on a scatterplot. These outliers may carry important information or indicate data quality issues.
- **Pattern Recognition:** Patterns such as clusters, curves, or groupings can be detected through scatterplots. This assisted us in identifying connections.
- **Variable Selection:** The evaluation of our independent variables was aided by scatterplots.

We can understand from **Figure 11** that there is an upward trend over time, indicating that the US GDP has grown significantly over the past two centuries. The US GDP reached $28 Trillion USD in 2023, up from $18 Trillion USD in 2003. This growth can be attributed to a number of factors, including population growth, technological advancement, and increased productivity.

Line Plot for Time Series Data

The complex patterns and trends buried in our time series data were largely revealed through the use of line plots. Gaining a better understanding of how different variables changed over time was the main goal. We carefully created these line plots, setting 'DATE' as the x-variable to represent the time progression chronologically, in order to accomplish this. A carefully chosen selection of parameters, including "open," "high," "low," and "close," were included in the y-variables. Line plots, a fundamental tool in time series analysis, hold the capacity to unveil a wealth of information. The curves and trajectories of each variable over time were visually depicted, allowing for intuitive recognition of fluctuations, periodic oscillations, or any significant deviations. By employingthis graphical approach, we obtained a holistic visualization of how each variable evolved, thus facilitating the identification of potential seasonality, trends, or irregularities. Such an exploration is crucial in various domains, particularly in financial analysis, and stress testing where open, high, low, and close prices provide essential insights for decision-making. The line plots presented a dynamic storyline of the data's temporal behavior, equipping us with a deeper comprehension of the relationships.

Correlation Analysis

```
                                                   WTI Prices  \
WTI Prices                                           1.000000
Crude Opil production Million Barrels/day            0.295393
US GDP (in Trillon USD)                              0.810654
Steel rebar futures price per tonne in Cinese Yuan   0.298170
U.S. Percent Utilization of Refinery Operable C...   0.657734
Opec Disruption                                     -0.684226
Non-OPEC Disruption                                 -0.064072
ISM Index                                            0.258957
Kilian Index                                         0.459744
Global stock                                        -0.096158

                                                   Crude Opil production Million Barrels/day  \
WTI Prices                                                                          0.295393
Crude Opil production Million Barrels/day                                           1.000000
US GDP (in Trillon USD)                                                             0.262597
Steel rebar futures price per tonne in Cinese Yuan                                 -0.162858
U.S. Percent Utilization of Refinery Operable C...                                  0.637611
Opec Disruption                                                                    -0.254586
Non-OPEC Disruption                                                                -0.570563
ISM Index                                                                          -0.270522
Kilian Index                                                                       -0.241762
Global stock                                                                       -0.179818
```

The "Correlation Analysis" is a crucial step that involved the computation of a correlation matrix. This matrix provided a numerical representation of how variables in our dataset were related to one another. It revealed not only whether these relationships existed but also the strength and direction of theseassociations. **Fig.12**, Shows the correlation between variables. Correlation coefficients, typically ranging from -1 to 1, indicated the degree of correlation. A value close to 1 indicated a strong positive correlation, implyingthat as one variable increased, the other also increased. Conversely, a value close to -1 indicated a strong negative correlation, where as one variable increased, the other decreased. There may be little to no linear correlation when the value is close to 0. We were able to understand how the variables were related to one another by displaying the correlation matrix as a table or heat map. With the help of this realization, we were able to determine which variables had a tendency to move in opposite directions when they were negatively correlated, and which way they were positively correlated. This

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

knowledge was essential for feature selection because highly correlated independent variables have the potential to cause multicollinearity in regression models, which could have an impact on the interpretability and functionality of the model. As such, one of the most important components of our methodology for understanding the intricate relationships in our dataset was the correlation analysis.

Heat – map Visualization

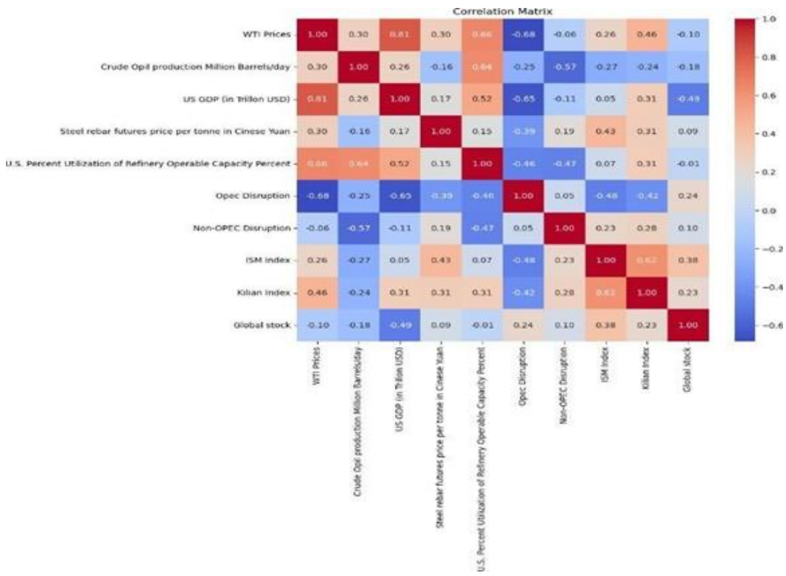

Figure 13: Correlation matrix to determine the relationship between variables in the dataset

There may be little to no linear correlation when the value is close to 0. We were able to understand how the variables were related to one another by displaying the correlation matrix as a table or heat map. With the help of this realization, we were able to determine which variables had a tendency to move in opposite directions when they were negatively correlated, and which way they were positively correlated. This knowledge was essential for feature selection because highly correlated independent variables have the potential to cause multicollinearity in regression models, which could have an impact on the interpretability and functionality of the model. As such, one of the most important components of our methodology for understanding the intricate relationships in our dataset was the correlation analysis.

Highlighting Patterns: The heat map made patterns in the data easier to see. In the visual representation, clusters of variables with positive or negative correlations stood out as unique patterns. It was possible to quickly grasp the interactions between subsets of variables thanks to this visual summary.

Data exploration: The heat-map served as a useful exploratory instrument, offering a foundation for more in-depth examinations. It inspired more research into relevant relationships and guided later modeling and research choices.

Statistical Analysis

VIF (Variance Inflation Factor)

	Variable	VIF
0	const	2630.622201
1	Crude Opil production Million Barrels/day	3.150261
2	US GDP (in Trillon USD)	2.441408
3	Steel rebar futures price per tonne in Cinese ...	1.455521
4	U.S. Percent Utilization of Refinery Operable ...	3.788466
5	Opec Disruption	3.206415
6	Non-OPEC Disruption	1.834663
7	ISM Index	2.637412
8	Kilian Index	2.837096

Figure 14: VIF calculated between each variable in the dataset

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

It is crucial to comprehend how the dependent (y) and independent (X) variables interact when performing regression analysis. When independent variables in a regression model show strong correlations or predictability among themselves, it is known as multicollinearity, and it can make the model difficult to interpret and use accurately. In this situation, the Variance Inflation Factor (VIF) is a potent instrument that provides insightful information about the correlations between these independent variables. Evaluating multicollinearity's existence and degree is the main goal of VIF. These are the main goals that this statistical method achieves:

Finding Multicollinearity: We can systematically ascertain whether independent variables in a regression model are correlated by using the VIF. High multicollinearity indicates that one or more variables can be predicted with the help of other variables, which may result in erratic coefficient estimates and impair the model's capacity to identify the particular impact of each variable on the dependent element. A VIF execution is shown in Fig. 14.

Measuring Multicollinearity's Extent: The Variance Inflation Factor (VIF) offers a numerical assessment of the degree of multicollinearity. Each independent variable's VIF score indicates the proportion of its variance that can be accounted for by the other independent variables in the model. Significant multicollinearity is indicated by a high VIF score, usually greater than 10.

Finding Problematic Variables: We can identify the precise independent variables that have the biggest impact on multicollinearity by using VIF analysis. This understanding helps us decide whether to keep these variables in the model, change them, or eliminate them in order to lessen the multicollinearity problem.

Regression Analysis

We used multiple regression analysis in an effort to gain a thorough understanding of the dataset and the variables influencing the behavior of the dependent variable. This method was an effective means of revealing the complex interactions that exist between the independent and dependent variables. Gaining understanding of the ways in which different attributes or factors, represented by the independent variables, impacted the behavior of the dependent variable was the main goal of this process. *The following steps encapsulate our regression analysis:* Multiple Regression Models: To explore the multivariate relationships in the dataset, multiple regression models were carefully constructed. These models took into account multiple independent variables at once, acknowledging that real-world phenomena are usually impacted by a wide range of variables. We were able to take into consideration how different attributes combined to affect the dependent variable. The multiple regression between each variable is shown in Fig. 15.

Feature Matrix and Target Vector: We carefully arranged our data into a feature matrix (X) and a target vector (y) to make the regression analysis easier. The features that we thought might have some explanatory powers were represented by the independent variables in the feature matrix (X). The dependent variable functioned as the focal point and was contained by the target vector (y). We were able to quantify and examine the connections between the independent and dependent variables thanks to this distinct demarcation.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             WTI Prices   R-squared:                       0.640
Model:                            OLS   Adj. R-squared:                  0.601
Method:                 Least Squares   F-statistic:                     16.69
Date:                Mon, 16 Oct 2023   Prob (F-statistic):           1.87e-09
Time:                        02:16:17   Log-Likelihood:                -209.18
No. Observations:                  53   AIC:                             430.4
Df Residuals:                      47   BIC:                             442.2
Df Model:                           5
Covariance Type:            nonrobust
==================================================================================================================
                                                      coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------------------------
const                                              10.5511     57.222      0.184      0.855    -104.565     125.667
Crude Opil production Million Barrels/day          -0.6324      0.885     -0.715      0.478      -2.412       1.148
Steel rebar futures price per tonne in Cinese Yuan  0.0001      0.003      0.050      0.960      -0.005       0.006
U.S. Percent Utilization of Refinery Operable Capacity Percent 1.6212 0.489 3.316 0.002 0.638 2.605
Opec Disruption                                   -14.5395      3.727     -3.901      0.000     -22.037      -7.042
Kilian Index                                        0.0364      0.048      0.752      0.456      -0.061       0.134
==============================================================================
Omnibus:                        5.186   Durbin-Watson:                   0.632
Prob(Omnibus):                  0.075   Jarque-Bera (JB):                4.580
Skew:                           0.718   Prob(JB):                        0.101
Kurtosis:                       3.117   Cond. No.                     1.30e+05
==============================================================================
```

Figure 15: Multiple regression between each variable to understand relationship between dependent and independent variables

This regression analysis was significant in two ways: Quantitative insights: Using multiple regression modeling, we were able to measure each independent variable's effect on the dependent variable quantitatively by utilizing statistical analysis. As a result, we were able to give each variable a coefficient that represented the strength and direction of its influence. Testing Hypotheses: Specific hypotheses regarding the ways in which the independent variables contributed to variations in the dependent variable were assessed through the use of regression analysis. The significance and applicability of these variables could be determined with the help of statistical tests and inference, which are supported by data.

Scatter Plot for Regression

The scatter plot shows a positive correlation between WTI prices and steel rebar futures prices, meaning that as WTI prices increase, steel rebar futures prices also tend to increase. This is because steel is a key input in the production of oil and gas, and rising oil prices can lead to higher costs for steel producers. Additionally, rising oil prices can lead to increased economic activity, which can boost demand for steel. **Figure 16** also shows that there is a lot of variation in the relationship between WTI prices and steel rebar futures prices. This is because there are a number of other factors that can affect steel prices, such as global economic conditions and government policies.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

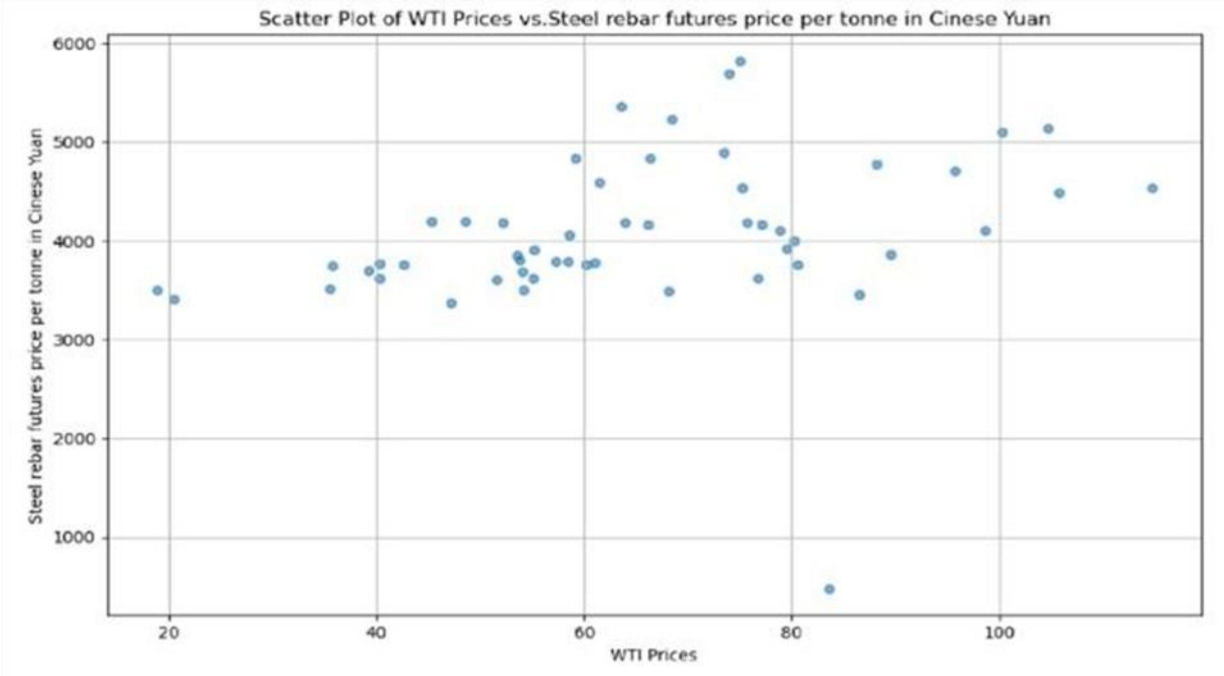

Figure 16: Scatterplot between WTI prices and Steel rebar future prices

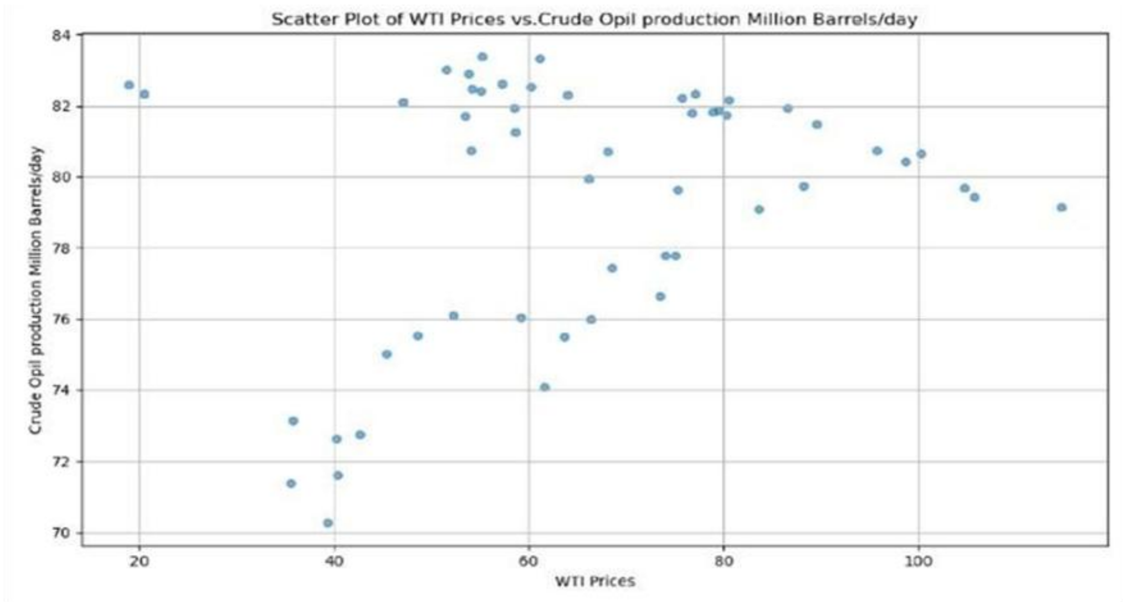

Figure 17: Scatterplot between WTI price and crude oil production

Figure 17 shows a positive correlation between WTI prices and crude oil production million barrels/day, meaning that as WTI prices increase, crude oil production million barrels/day tends to increase. This is because crude oil is a key input in the production of oil and gas, and rising WTI prices can lead to increased exploration and production of crude oil. The scatter plot also shows that there is a lot of variation in the relationship between WTI prices and crude oil production million barrels/day. This is because there are a number of other factors that can affect crude oil production, such as global economic conditions, geopolitical factors, and technological advancements.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

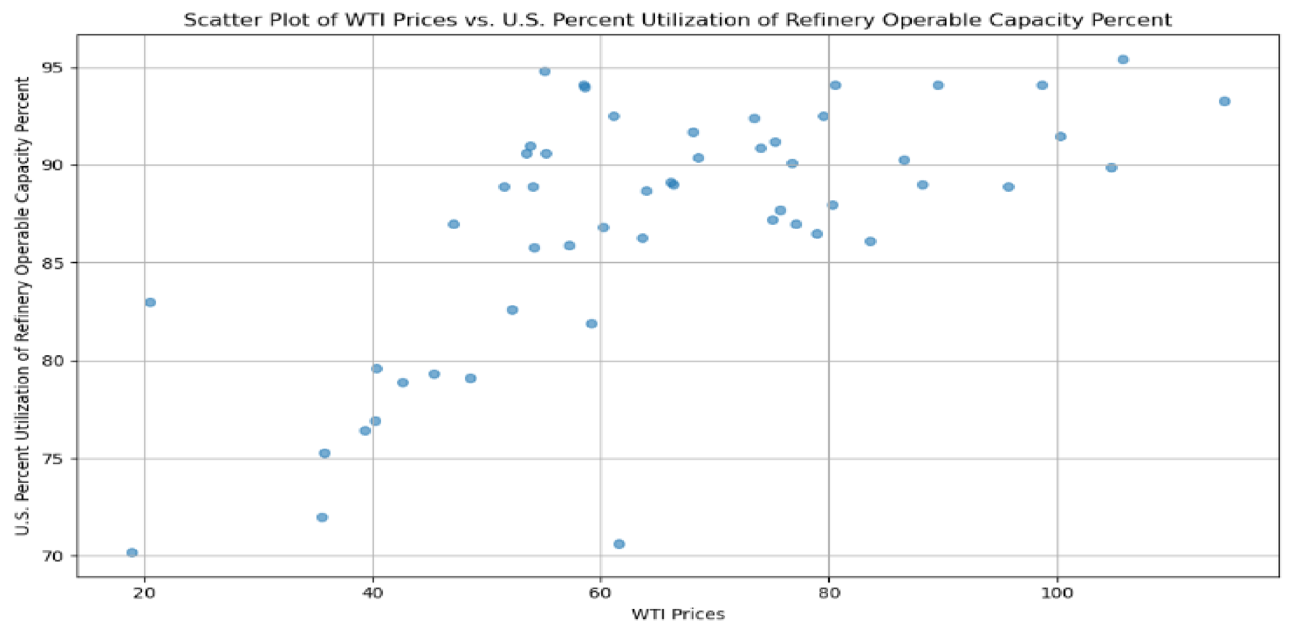

Figure 18: Scatterplot between WTI prices and US percent utilization of refinery operable capacity

We created scatter plots in this step (Fig. 18) to show the relationships between the dependent variable, "WTI Crude," and the other independent variables. In order to interpret the results of the regression model, it is crucial to investigate relationships and patterns between the variables, and these scatter plots provided as a fundamental tool for doing so. We can see how "WTI Crude" relates to each independent variable separately thanks to the scatter plots (Figs. 16 and 17). Plotting the data points allowed us to visually evaluate the distribution of the data overall, possible outliers, and linear or nonlinear associations. The underlying structure of the data and the formation of preliminary hypotheses regarding the relationships between variables were greatly aided by these visualizations. **Figure 18** shows a positive correlation between WTI prices and US percent utilization of refinery operable capacity percent, meaning that as WTI prices increase, US percent utilization of refinery operable capacity percent tends to increase. This is because when WTI prices are high, refineries are more likely to operate at full capacity in order to produce more gasoline and other petroleum products [18]. **Figure 18** also shows that there is a lot of variation in the relationship between WTI prices and US percent utilization of refinery operable capacity percent. This is because there are a number of other factors that can affect refinery utilization rates, such as inventory levels, maintenance schedules, and unexpected events such as hurricanes or refinery figures [19].

Estimation of Model

We As a crucial first step in our analysis, we estimated the statistical model. Finding the regression coefficients, which are essential for comprehending the relationship between the independent and dependent variables, was the main goal of this stage, as shown in Fig. 19. These coefficients offer crucial information by clarifying the direction and strength of each independent variable's influence on the dependent variable. The determined coefficients showed the relationship between changes in an independent variable's unit and changes in the dependent variable. Comprehending these correlations is essential for deriving conclusions, forecasting, and analyzing the model's output. Several parts of our analysis were based on the results of this estimation process. Several parts of our analysis were based on the results of this estimation process. It gave us more insight into the underlying dynamics of the dataset in addition to enabling us to forecast the dependent variable based on the values of the independent variables.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

```
Crude Opil production Million Barrels/day:
     Coefficient: -0.6324
     Standard Error: 0.885
     t-value: -0.7145762711864406
     p-value: 0.47840522644134653

Steel rebar futures price per tonne in Cinese Yuan:
     Coefficient: 0.0001
     Standard Error: 0.003
     t-value: 0.03333333333333333
     p-value: 0.9735499331106836

U.S. Percent Utilization of Refinery Operable Capacity Percent:
     Coefficient: 1.6212
     Standard Error: 0.489
     t-value: 3.3153374233128834
     p-value: 0.0017696862248344214

Opec Disruption:
     Coefficient: -14.5395
     Standard Error: 3.727
     t-value: -3.9011269117252483
     p-value: 0.00030365409492683O7
```

Figure 19: Regression coefficients for each variable

Hypothesis Testing

The phase of hypothesis testing was essential to our process. Its main objective was to evaluate the regression coefficients in our model for statistical significance. At this point, we used statistical methods that required figuring out two-tailed p-values and t-values. We employed these values to assess the reliability of our conjectures concerning the correlations between the independent and dependent variables.

The core objectives of hypothesis testing were twofold:

Coefficient Significance: One of the primary goals was to determine whether individual regression coefficients (representing the effect of each independent variable on the dependent variable) were statistically significant. By doing so,we could ascertain which variables had a substantial impact on the model's outcome.

Model Assessment: In the broader context, hypothesis testing contributed to the overall assessment of our regression model's validity and effectiveness. It allowed us to understand which variables should be retained in the model and which could be omitted, streamlining the model's complexity while preservingits predictive power.

Regression Equation

In the context of our analysis, the regression equation served as a fundamental outcome derived from the estimated coefficients of the multiple regression model. This equation stands as a mathematical model that plays a central role in predicting the dependent variable concerning variations in the independent variables. It encapsulates the linear relationship between these variables and is expressed in a mathematical form. The regression equation's essence lies in its predictive capability. By inputting specific values for the independent variables, this equation generates an estimated value for the dependent variable. In essence, it quantifies how changes in the independent variables contribute to changes in the dependent variable, facilitating predictions and understanding the effect of each independent variable on the outcome.

Probability Distribution Identification

Figure 20 shows that the KS test statistic is 1.0 and the p-value is 0.0. This means that the probability of obtaining a test statistic of 1.0 or greater under the null hypothesis that the data is normally distributed is 0.0. Therefore, we can reject the null hypothesis and conclude that the data is not normally distributed.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

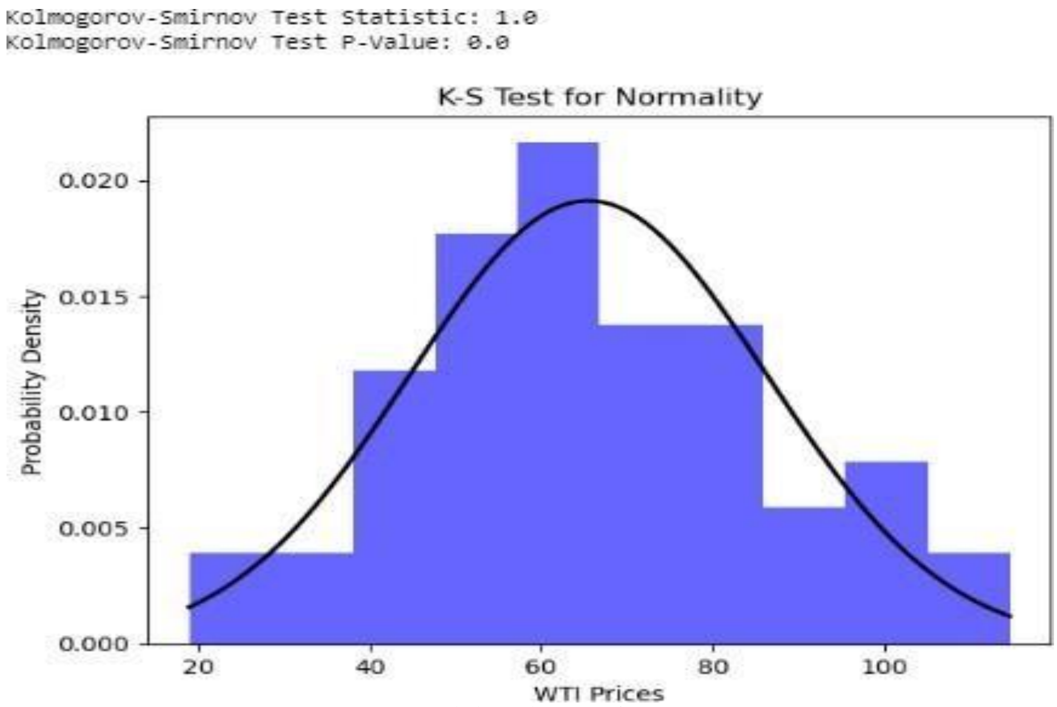

Figure 20: Normal distribution for WTI prices

The slope in the data is likely due to the presence of outliers or non-normality in the data. Outliers can cause the KS test to reject the null hypothesis even if the data is approximately normally distributed. Non-normality in the data can also cause the KS test to reject the null hypothesis.

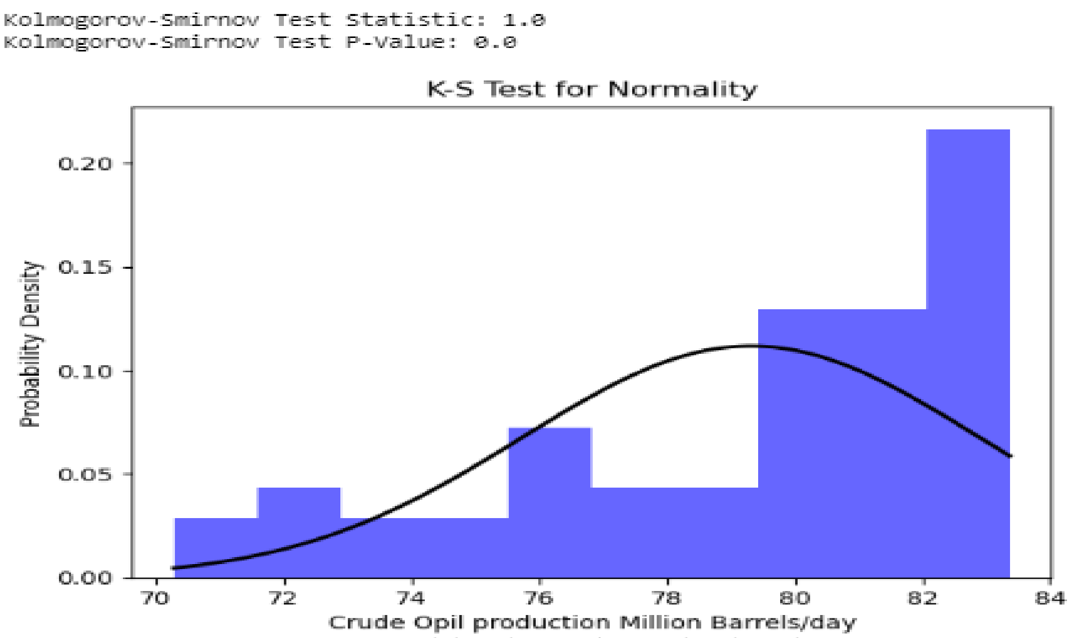

Figure 21: Normal distribution for Crude oil production

Figure 21 explains that the distribution is right-skewed, with a median production of 74 million barrels per day and a maximum production of 84 million barrels per day. The Kolmogorov-Smirnov test statistic is 1.0 and the p-value is 0.0, which indicates that the distribution is significantly different from a normal distribution. This distribution can be explained by the fact that crude oil production is concentrated in a few key regions of the United States, such as the Permian Basin and the Gulf of Mexico. These regions have very high production rates, which skew the distribution to the right.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

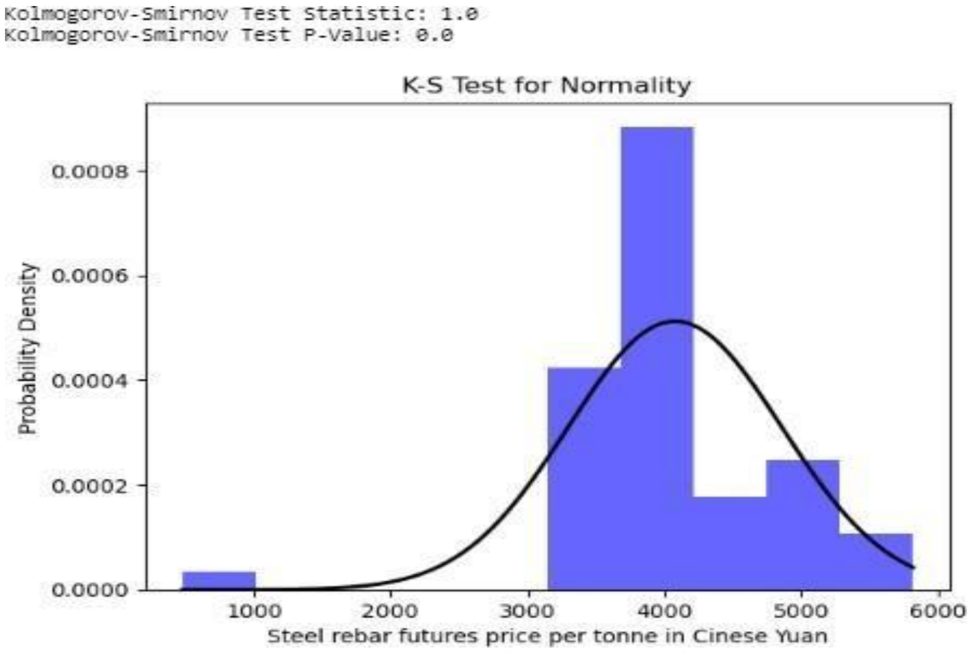

Figure 22: Normal distribution for Steel rebar future price

In Figure 22 the distribution is skewed to the right, which means that there are more futures contracts at lower prices than at higher prices. The median price is 3,980 yuan per tonne, and the maximum price is 6,000 yuan per tonne. The Kolmogorov-Smirnov (KS) test statistic is 1.0 and the p-value is e.e, which means that the distribution is significantly different from a normal distribution. This is likely due to the fact that steel rebar futures prices are affected by a number of factors, including supply and demand, economic growth, and government policies.

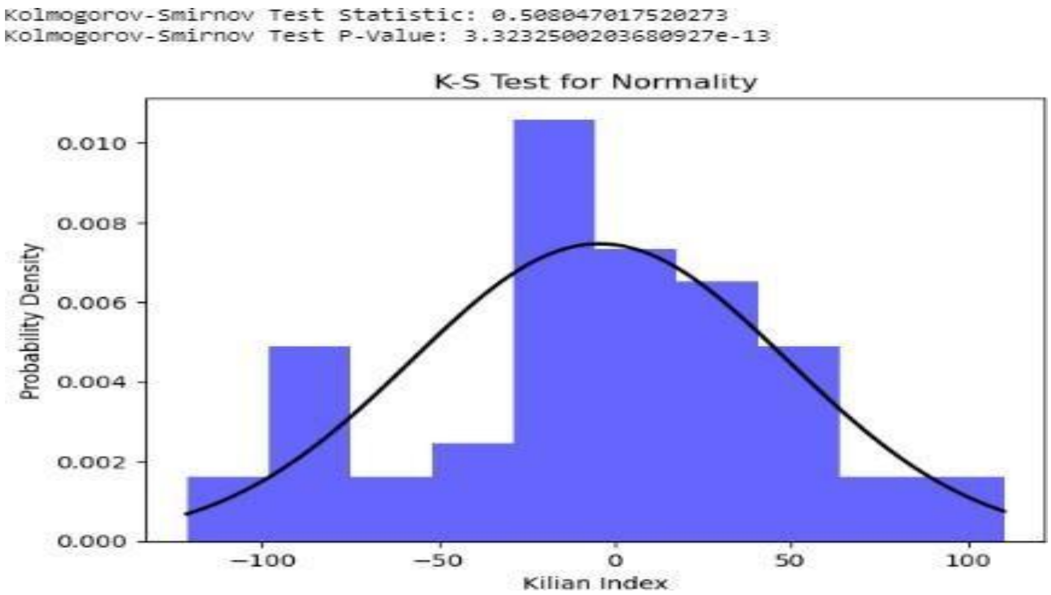

Figure 23: Normal distribution for Killian Index

We employed the Kolmogorov-Smirnov test (Fig.20, Fig.21, Fig.22, and Fig.23) to evaluate the suitability of a specific probability distribution for the dataset. The main goal was to determine how closely the observed data fit the parameters of a chi-squared distribution. We were able to determine whether the selected distribution was appropriate thanks to this statistical test, which also helped us comprehend how the data adhered to or deviated from the expected distribution pattern. It was an essential step in confirming the underlying hypotheses and offered insightful information about the statistical behavior of the data. Figure 23 shows a graph of a normal distribution with a slope of 0.5. This means that the distribution is skewed to the right, with more values on the right-hand side

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

of the distribution than on the left-hand side. The distribution also has a relatively long tail on the right-hand side, which means that there is a small chance of observing very high values.

Calculation of Quartiles

```
1st Quartile (Q1): 78.99127535207676
3rd Quartile (Q3): 79.63463453175795

1st Quartile (Q1): 4008.606408302699
3rd Quartile (Q3): 4148.152554147412

1st Quartile (Q1): 86.44906162393l3
3rd Quartile (Q3): 87.5899472954512

1st Quartile (Q1): 2.446277981136894
3rd Quartile (Q3): 2.560408315645384
```

Figure 24: Calculation of 1st and 3rd quartiles

In order to obtain a better understanding of the dataset's distribution, we computed the first quartile (Q1) and the third quartile (Q3) for Fig. 24 during this methodology phase. The quartiles offered helpful benchmarks for comprehending the distribution and properties of the variables being studied. Determining Q1 and Q3 allowed us to gain the following advantages: Distribution Assessment: One important step in determining the dataset's distribution was to compute the quartiles. It enabled us to pinpoint the locations in the dataset where specific percentages of the data—25% and 75%—were concentrated. Quartiles were crucial instruments for the identification of outliers. A comparison with quartile values made it simple to identify outliers, or data points that differ significantly from the majority of the data [25]. Data points falling below Q1 or above Q3 could be potential outliers warranting further investigation.

Data Segmentation: Understanding quartiles also facilitated data segmentation. It served as a foundation for grouping data points into various quartile-based ranges or groups, which is particularly helpful when examining how a variable behaves across various dataset segments.

Moving Average

```
Minimum value in 'low' category: 83.59704010250022
Maximum value in 'high' category: 90.48871211628509
Average for 'normal' category: 87.01972287943714

Minimum value in 'low' category: -33.86842408175326
Maximum value in 'high' category: 26.277912771950227
Average for 'normal' category: -4.282385411319961

Minimum value in 'low' category: 2.141479860768024
Maximum value in 'high' category: 2.874889222941288
Average for 'normal' category: 2.503145087387l455

Minimum value in 'low' category: 77.3065858483752
Maximum value in 'high' category: 81.22557871018509
Average for 'normal' category: 79.31320362354097
```

Figure 25: Three point moving average

We used a 3-point moving average on all of the dataset's variables. This operation's main goal was to smooth the data so that underlying trends and patterns could be easier to identify. By implementing a moving average, we reduced the impact of short-term fluctuations or noise in the data, making it easier to discern meaningful trends and variations [24]. In **Fig.25**, This process helped in revealing the broader patterns andmovements that might not have been apparent in the raw data. In essence, the application of a 3-point moving average served as a data preprocessing technique that enhanced the dataset's suitability for time series analysis, ultimately enabling us to extract more accurate and actionable insights from the data [20].

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Natural Logarithm Transformation

	DATE	WTI Prices	Crude Opil production Million Barrels/day \
52	2019-01	53.79	82.91
51	2019-02	57.22	82.61
50	2019-03	60.14	82.54
49	2019-04	63.91	82.30
48	2019-05	53.50	81.71
47	2019-06	58.47	81.93
46	2019-07	58.58	81.26
45	2019-08	55.10	82.40
44	2019-09	54.07	80.73
43	2019-10	54.18	82.47

	Steel rebar futures price per tonne in Cinese Yuan \
52	3803.9
51	3787.9
50	3757.6
49	4181.8
48	3848.5
47	3787.9
46	4060.6
45	3621.2
44	3681.8
43	3507.6

	U.S. Percent Utilization of Refinery Operable Capacity Percent \
52	91.0
51	85.9
50	86.8
49	88.7
48	90.6
47	94.1
46	94.0
45	94.8
44	88.9
43	85.8

	Opec Disruption	Kilian Index
52	2.50	-37.381736
51	2.57	-89.925034
50	2.19	-82.308332
49	2.18	-69.718519
48	2.50	-40.297801
47	2.55	-27.574185
46	2.65	18.939382
45	2.65	24.777347
44	3.96	37.672665
43	2.65	16.452426

Figure 26: Natural log transformation to handle exponential values

As part of our data preparation process, we applied a natural logarithm transformation to all fields (columns) within the dataset. This transformation played a crucial role in ensuring that the data met the fundamental assumptions required for regression analysis, with a particular focus on datasets exhibiting exponential growth characteristics, this is represented in Fig.26. The natural logarithm transformation is widely used in statistical modeling to address two primary issues:

Variance Stabilization: Data with exponential growth tendencies often exhibitincreasing variance as the values grow larger. This can lead to heteroscedasticity, a violation of the homoscedasticity assumption in regression analysis. By applying the natural logarithm transformation, we mitigated this issue, making the variance more consistent across the range of values. This, in turn, aids in producing reliable and interpretable regression models.

Linearity Enhancement: Regression models, including linear regression, assume a linear relationship between the independent and dependent variables.However, in practice, some variables may not exhibit linear relationships in their raw form [23]. The natural logarithm transformation can linearize such relationships, making the data more amenable to regression analysis.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

RESULTS ANALYSIS

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             WTI Prices   R-squared:                       0.831
Model:                            OLS   Adj. R-squared:                  0.820
Method:                 Least Squares   F-statistic:                     76.94
Date:                Tue, 17 Oct 2023   Prob (F-statistic):           3.74e-18
Time:                        00:36:31   Log-Likelihood:                 29.024
No. Observations:                  51   AIC:                            -50.05
Df Residuals:                      47   BIC:                            -42.32
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================================================
===
                                                          coef    std err          t      P>|t|      [0.025      0.9
75]
--------------------------------------------------------------------------------------------------------------
---
const                                                  -1.6043      2.142     -0.749      0.458      -5.914       2.
705
Crude Opil production Million Barrels/day              -1.7648      0.693     -2.548      0.014      -3.158      -0.
371
U.S. Percent Utilization of Refinery Operable Capacity Percent     3.2896      0.492      6.689      0.000       2.300       4.
279
Opec Disruption                                        -0.9906      0.145     -6.849      0.000      -1.282      -0.
700
==============================================================================
Omnibus:                        0.048   Durbin-Watson:                   0.317
Prob(Omnibus):                  0.976   Jarque-Bera (JB):                0.066
Skew:                           0.047   Prob(JB):                        0.968
Kurtosis:                       2.852   Cond. No.                         709.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 27: Regression model performance

In **Fig.27**, The analysis of the regression model's performance is essential to understand how well it fits the data. In this section, we present the results of the model bycomparing the actual and estimated values through scatterplots and analyzing the confusion matrix for classification tasks. From **Fig.27** We can see that the R-squared value is 88.31%, which indicates that the model explains a high proportion of the variation in WTI crude oil prices. The F-statistic is also highly significant, which means that the model is a good fit for the data. The coefficient for crude oil production is negative, which indicates that an increase in crude oil production leads to a decrease in WTI crude oil prices. This is because an increase in supply leads to downward pressure on prices. The coefficient for U.S. percent utilization of refinery operable capacity is positive, which indicates that an increase in refinery utilization leads to an increase in WTI crude oil prices. This is because an increase in demand for crude oil leads to upward pressure on prices. The coefficient for OPEC disruption is negative, which indicates that an increase in OPEC disruption leads to a decrease in WTI crude oil prices. This is because OPEC is a major producer of crude oil, and a disruption to OPEC production can lead to a decrease in global supply and an increase in prices.

Actual vs Estimated Scatterplot

Making scatterplots that show the actual values versus the estimated values is a popular method in regression analysis to assess the effectiveness of the model. These scatterplots in Figure 28 give an idea of how closely the model predictions match the actual data. The "Actual vs. Estimated Scatterplot" shows how the regression model's predicted values and true values relate to one another. We can evaluate the accuracy of the model visually by charting these values. A diagonal line should ideally be closely aligned by the points to show that the model's predictions and the actual data points are nearly identical [21], these graphs in Figure 28 also shows a confidence interval for the estimated prices. The confidence interval is a range of values that is likely to contain the true value of the parameter being estimated. In this case, the confidence interval is for the estimated WTI prices.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

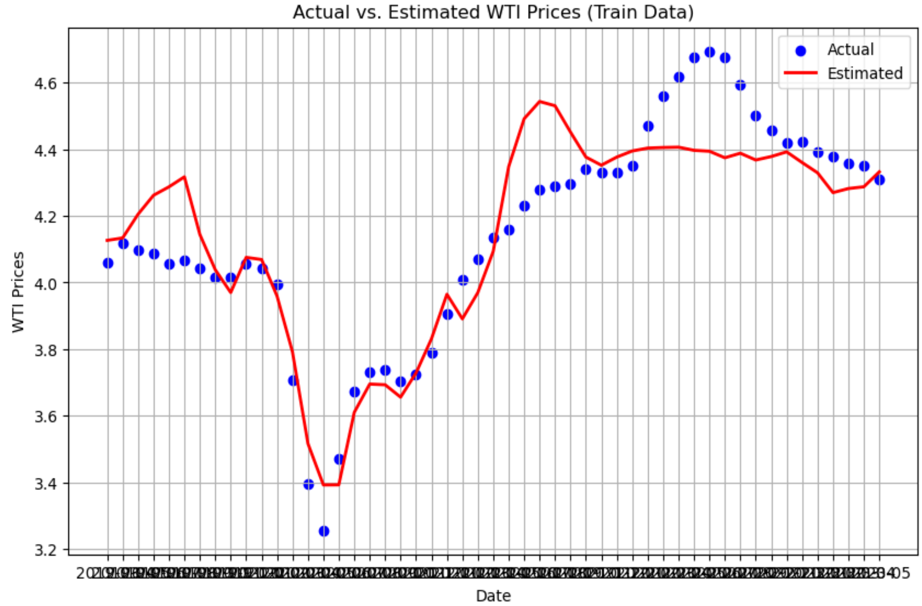

Figure 28: Scatterplot between Actual and Estimated value

The confidence interval is wider in some areas than in others. This means that the model is more uncertain about the estimated prices in those areas [22].

Confusion Matrix Analysis

```
Confusion Matrix:
[[51]]
Precision: 1.00
Recall: 1.00
F1 Score: 1.00
Accuracy: 1.00
```

Figure 29: confusion matrix

Analyzing the confusion matrix is necessary to evaluate the model's performance for classification tasks in Fig. 29. A tabular representation of a classification model's performance that provides a thorough breakdown of true positive, true negative, false positive, and false negative predictions is called a confusion matrix. We analyze the confusion matrix to determine how well the model performs. We look at metrics like accuracy, recall, F1-score, and precision. These metrics shed light on how well the model classifies instances into the appropriate categories. We can evaluate the classification model's advantages and disadvantages and decide how best to apply it in practice by looking at the confusion matrix.

CONCLUSION

In Conclusion, this analysis for assessing dependencies provides valuable insights into crude oil markets, the impact of financialization on these markets, and the broader landscape of risk assessment in the energy sector. The methodologies and principles outlined underscore the importance of robust risk management practices in safeguarding energy security. The interplay of market risk, macroeconomic scenarios, and supply and demand factors reveals the intricate web of challenges that energy companies, policymakers, and researchers must navigate.

Furthermore, the evolving nature of crude oil markets and the impact of financialization on these markets shed light on the need for continuous monitoring and adaptation. The studies on diversification benefits, extreme risk modeling, and the common factors influencing crude oil prices contribute to a more comprehensive understanding of these complex markets. In this context, the works emphasize the importance of staying current with the latest methodologies and market trends, adapting to changing market dynamics, and ensuring that energy companies are well-prepared to withstand various stress scenarios. As a result, the knowledge and insights gained from this body

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

of work are crucial for practitioners, policymakers, and researchers seeking to navigate through crude oil market dynamics and risk management

REFRENCES

[1] Mazing, C. (2014). Introducing a holistic approach to stress testing. Source Details.
[2] Bocchio, C. (2015). Stressed scenarios and linkages to market risk instruments. Source Details.
[3] Briand, R. (Year). Scenarios, stress tests, and strategies for 2016. Source Details.
[4] European System of Financial Supervision. (2023). Macro financial scenario for the 2023 EU-wide banking sector stress test. Source Details.
[5] Adrian, T., Morsink, J., & Schumacher, L. (2020). "Stress Testing at the IMF." Departmental Paper No. 20/04, Basel Committee for Banking Supervision.
[6] IMF Monetary and Capital Markets. (2015). Assessing stress from oil price shocks. Source Details.
[7] Bank of England. (2023). Key elements of the 2023 CCP supervisory stress test. Source Details.
[8] Paraschiv, F., Reese, S. M., & Skjelstad, M. R. (2019). Portfolio stress testing applied to commodity futures. Source Details.
[9] Glück, Z., & Glück, T. (2015). Financialization in commodity markets: A passing trend or the new normal? Journal of Banking & Finance, 60, 93-111.
[10] Aepli, M. (2011). On the design of stress tests. Master's thesis, University of St. Gallen.
[11] Aepli, M., Füss, R., Henriksen, T. E. S., & Paraschiv, F. (2017). Modeling the multivariate dynamic dependence structure of commodity futures portfolios. Journal of Commodity Markets, 6, 66-87.
[12] Alexander, C. (2008a). Market risk analysis: Practical financial econometrics, vol. 2. Wiley.
[13] Alexander, C. (2008b). Market risk analysis: Value at risk models, vol. 4. Wiley.
[14] Baffes, J., Kose, M. A., Ohnsorge, F., & Stocker, M. (2015). The great plunge in oil prices: Causes, consequences, and policy responses (policy research note). World Bank Group.
[15] Basak, S., & Pavlova, A. (2016). A model of financialization of commodities. Journal of Finance, 71(4), 1511-1556.
[16] Basel Committee on Banking Supervision. (2006). International convergence of capital measurement and capital standards. Technical report, Bank for International Settlements.
[17] Basel Committee on Banking Supervision. (2009). Principles for sound stress testing practices and s upervision. Technical report, Bank for International Settlements.
[18] Basel Committee on Banking Supervision. (2017). Supervisory and bank stress testing: Range of practices. Technical report, Bank for International Settlements.
[19] Basel Committee on Banking Supervision. (2018). Stress testing principles. Technical report, Bank for International Settlements. ISBN 978-92-9259-207-3 (online).
[20] Baudino, P., Goetschmann, R., Henry, J., Taniguchi, K., & Zhu, W. (2018). FSI insights on policy implementation no. 12: Stress-testing banks--a comparative analysis. Technical report, Financial Stability Institute. ISSN 2522-2481 (print).
[21] Bhardwaj, G., Gorton, G., & Rouwenhorst, G. K. (2015). Facts and fantasies about commodity futures ten years later (working paper). National Bureau of Economic Research, Cambridge.
[22] Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control. Wiley, Hoboken.
[23] Cheng, I. H., & Xiong, W. (2014). Financialization of commodity markets. Annual Review of Financial Economics, 6(1), 419-441.
[24] Cheung, C. S., & Miu, P. (2010). Diversification benefits of commodity futures. Journal of International Financial Markets, Institutions & Money, 20(5), 451-474.
[25] Daskalaki, C., & Skiadopoulos, G. (2011). Should investors include commodities in their portfolios after all? New evidence. Journal of Banking & Finance, 35(10).