

# ***Multiple Stock Prediction Based on Linear and Non-linear Machine Learning Regression Methods***

**Shuqi Chen<sup>1,a,\*</sup>**

<sup>1</sup>*Economics and Statistics major, University of Toronto, 27 King's College Cir, Toronto, ON, Canada*

*a. shuqichen.chen@mail.utoronto.ca*

*\*corresponding author*

**Abstract:** In contemporary times, the pressing issue of global environmental pollution has prompted the exploration of alternative energy sources by various industries, aiming to mitigate the adverse environmental impacts caused by traditional energy production. Correspondingly, investors in the financial market have increasingly redirected their capital towards the new energy sector. Within this context, the present research endeavors to employ machine learning techniques for the prediction of Tesla's stock price. This study leverages multiple linear regression, polynomial regression, and lag models to construct models based on the datasets of TSLA, MPC, and UNG stock prices spanning the period of 2019-2020. By discerning potential patterns among these variables, the objective is to anticipate the future trajectory of TSLA stock price. According to machine learning methods, Tesla's stock price can be predicted, and the daily price of Tesla is influenced by the opening price, high price, low price and trading volume of the stock on that day. In addition, the share prices of energy companies related to Tesla also have an impact on Tesla's share price on that day. Specifically, Tesla's stock price is influenced by Natural Gas Company (UNG), which has an opposite relationship. Although common sense economics says that the crude oil market will be closely related to the new energy market. However, the results of this study demonstrated that Tesla's stock price is less influenced by Crude Oil Company (MPC).

**Keywords:** TESLA stock price prediction, machine learning, non-linear regression

## **1. Introduction**

The new energy industry has garnered considerable attention over an extended period. With the rapid development of modern technology, people are focusing more attention on environmental protection issues and are increasingly aware of the negative impact of traditional fuels on the environment. The use of large amounts of traditional energy sources, such as crude oil and natural gas, can have irreversible negative effects on the environment [1]. In this context, the search for clean energy sources such as solar and wind energy has been focused on replacing traditional energy sources. The escalating demand for environmentally friendly and sustainable energy sources has led to a surge in investor interest in the prospective new energy market. Primarily, studying the new energy markets of the future can facilitate the identification of emerging technologies and trends that are conducive to reducing carbon emissions and promoting a cleaner environment. In addition, governments and the international community have invested a lot of support in projects to develop clean energy, and the

new energy market has great potential for future growth. Therefore, studying new energy technology trends can help companies stay at the forefront of innovation. And from an investor's perspective, it becomes imperative to forecast the future development of the new energy stock market, enabling investors to maximize their profits.

In recent years, with the development of artificial intelligence and computer science, many statistical and machine learning techniques have been widely applied to other fields. In the field of economics and finance, machine learning has shown significant usefulness. Economists need to process and analyze a lot of data, which, when combined with effective data models, helps economists to better organize and analyze historical data and draw conclusions. In the beginning, machine learning was only a small part of artificial intelligence and was not widely used. It was not until later that many studies started to combine econometrics with machine learning, catering to the era of big data development and making economics research more efficient [2]. For example, in the fields of environmental economics and energy, common data collection and research methods are not sufficient to reflect the significant modeling and complexity of large data sets, so it is necessary to use machine learning in the analysis of economic and financial problems.

In fact, stock prices are subjected to many factors, such as the daily trading volume of stocks, investors' expectations of the market, and policy conditions [3]. However, since stock prices are so volatile that it is challenging to collect a comprehensive dataset that affects them, this study chose a dataset from the traditional energy market that is inversely related to the new energy market and tried to study the relationship between stock prices of different energy sources. As a result of the energy revolution, the global energy use structure has undergone a reformation process from coal to oil to hydrocarbons and finally to new energy sources [1]. Given the nascent stage of research on new energy sources, the energy sector remains in a transitional phase. Consequently, an investigation into the oil market and other energy sources, such as natural gas can contribute to better understanding of the new energy market and find the relationship between these different types of energy sources.

## 2. Methodology

### 2.1. Dataset Collection and Description

The datasets utilized in this study are the stock prices of a few of the most representative companies in each energy sector from 04/01/2019 to 06/30/2021. Among them, TSLA represents the new energy sector (TESLA), Marathon Petroleum Corporation (MPC) represents the oil market, and United States Natural Gas Fund (UNG), represents the stock price of the natural gas market. For each data set, they have the same variables, namely date, open price, close price, high price, low price, and trading volume. These datasets were collected from Yahoo Finance official website. The stock price movement of new energy is related to the oil price [4], which in turn is directly reflected in the stock price of oil companies, therefore, this study selected the stock price dataset of MPC oil companies. In the modeling process, this study used the close price as the daily stock price for the calculation. Table 1 shows the features and corresponding explanations.

Table 1: The feature and corresponding explanation of the dataset.

Feature Name	Value
Date (Numerical)	The date corresponding to the stock data.
Open (Numerical)	The open price of daily stock.
High (Numerical)	The highest stock price each day.
Low (Numerical)	The lowest stock price each day.
Close (Numerical)	The close stock price each day.
Volume (Numerical)	Total daily stock trading volume.

### 2.1.1. Dataset Cleaning Process

Prior to analysis, a preliminary data cleansing process was conducted on the raw dataset. Null variables were identified and subsequently eliminated, resulting in a dataset comprising a total of 568 valid data points for all three companies. This limited the dataset to include only those days with complete information across all variables. Consequently, the subsequent task of merging and comparing the daily stock prices of the different companies was facilitated by the availability of this consolidated dataset.

### 2.1.2. Variables Introduction

The overall view of the data shows that Tesla stock price will move a lot between 2019 and 2021, and that UNG stock price will move very little and tend to be stable compared to it. While UNG share price is less volatile compared to TESLA and MPC, it remains more volatile when looking at UNG's share price in isolation from 2019 to 2020 because the benchmark share price is different and Tesla's unit price per share is much higher than UNG stock price. The summary table of these stock prices is shown in Table 2.

Table 2: Summary table for US natural gas, TSLA and MPC market prices from 2019/03/30 to 2021/06/30.

Variable	Minimum	Q1	Median	Q3	Maximum	IQR	Mean	Standard deviation
Natural Gas Close Price	8.48000	10.60000	12.85000	19.2325	23.8500	8.6325	14.58905	4.525665
TSLA Close Price	11.9313	19.2236	54.6096	191.569	294.3633	172.345	100.0323	88.39995
MPC Close Price	16.6200	37.2550	50.6650	58.5425	68.0200	21.2875	47.58060	12.34687

### 2.1.3. TESLA K Line

Looking at the K-line chart shown in Figure 1, Figure 2 and Figure 3. In February 2020, MPC stock price experienced a substantial and sharp decline, but after that, MPC overall stock price trended upward. As MPC stocks fell sharply, TESLA stocks also tended to fall to some extent at the same time. Comparing the stock price curves of TESLA and UNG, TESLA stock price is generally trending upward while UNG stock price is trending downward during the period 1029 to 2021. Moreover, their fluctuations are almost opposite.



Figure 1: Tesla stock K-Chart (from 2019/04/01 to 2021/06/29) (photo/picture credit: original).



Figure 2: MPC stock K-Chart (from 2019/04/01 to 2021/06/29) (photo/picture credit: original).

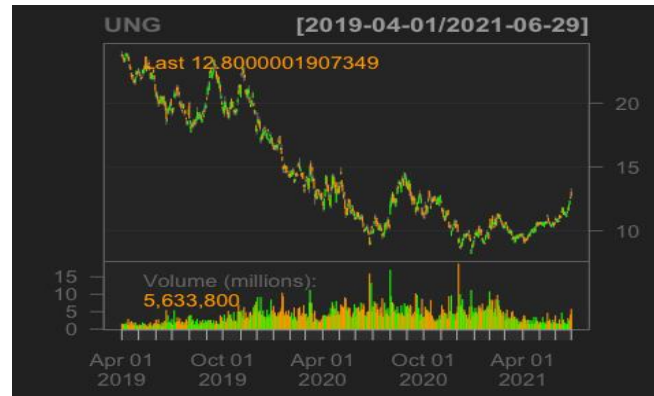


Figure 3: UNG stock K-Chart (from 2019/04/01 to 2021/06/29) (photo/picture credit: original).

## 2.2. Machine Learning Models

### 2.2.1. Multiple Linear Regression Model

For Tesla's own stock closing price prediction, other stock trading data for the day can be used. For this purpose, a multiple regression model was used to try to find the relationship between Tesla's daily stock closing price and its opening price, high price, low price, and trading volume for that day [5]. The multiple linear regression model used is shown in equation (1). Additionally, the explanation of each symbol is provided in Table 3.

$$Close = \beta_0 + \beta_1 * X_{high} + \beta_2 * X_{low} + \beta_3 * X_{open} + \beta_4 * X_{volume} + E \quad (1)$$

Table 3: The explanation of each symbol.

$E$	$E$ means error term.
$\beta_0$	$\beta_0$ represents the interception of daily stock close price.
$\beta_1$	$\beta_1$ is the coefficient of daily high price.
$\beta_2$	$\beta_2$ is the coefficient of daily low price.
$\beta_3$	$\beta_3$ is the coefficient of daily open price.
$\beta_4$	$\beta_4$ is the coefficient of daily stock trading volume.
$X_{high}$	$X_{high}$ is the daily highest stock price.
$X_{low}$	$X_{low}$ is daily lowest stock price.
$X_{open}$	$X_{open}$ is daily open price of stock.
$X_{volume}$	$X_{volume}$ is the total daily stock trading volume.

According to the stock K line plot above, it finds that Tesla and UNG daily stock prices move in opposite trends, then the study tries to establish a regression relationship between the three share prices. The formula is shown in equation (2). Additionally, the explanation of each symbol is provided in Table 4.

$$Close_{TSLA} = \beta_0 + \beta_1 X_{UNG} + \beta_2 X_{MPC} \quad (2)$$

Table 4: The explanation of each symbol.

$\beta_0$	$\beta_0$ is the interception of TESLA close price
$\beta_1$	$\beta_1$ is the coefficient of UNG close price
$\beta_2$	$\beta_2$ is the coefficient of MPC close price
$X_{UNG}$	$X_{UNG}$ represents the close price of UNG
$X_{MPC}$	$X_{MPC}$ represents the close price of MPC

### 2.2.2. Polynomial Regression Model Method

Considering that the complex relationships between stock close price and other data, the common linear regression models may not adequately capture the complexity of the relationship, however, some non-linear models may illustrate them better. Therefore, this study tried the polynomial regression model to find the relationships between TESLA's daily close price and other variables. To predict the daily close price of TESLA, I focus on TESLA's daily high price, low price and open price. The polynomial regression formula is:

$$Close_{TSLA} = \beta_0 + \beta_1 * X + \beta_2 * X^2 + \beta_3 * X^3 + \beta_4 * X^4 \quad (3)$$

$\beta_0, X$  represents the interception of daily stock close price and the independent variables (daily high/low/open price), respectively.

### 2.2.3. Lag Modeling Method

To better predict the future Tesla stock price, this study used the Lag model, adding a lag factor to the original regression model. In other words, the past stock price is used to predict the future stock price. It is believed that Tesla stock price is affected by its own lagged stock price as well as close prices of MPC and UNG the day before. Therefore, this study calculates the lag values of TESLA, MPC, and UNG daily stock prices and model them to illustrate the own stock price of Tesla. The final formula is shown in equation (4). Additionally, the explanation of each symbol is provided in Table 5.

$$Close_{TSLA} = \beta_0 + \beta_1 * lag(X_{TSLA}) + \beta_2 * lag(X_{UNG}) + \beta_3 * lag(X_{MPC}) \quad (4)$$

Table 5: The explanation of each symbol.

$\beta_0$	$\beta_0$ is the interception of TESLA close price
$\beta_1$	$\beta_1$ is the coefficient of TESLA lagged close price
$\beta_2$	$\beta_2$ is the coefficient of UNG lagged close price
$\beta_3$	$\beta_3$ is the coefficient of MPC lagged close price
$lag(X_{TESLA})$	$lag(X_{TESLA})$ represents the lagged close price of TESLA
$lag(X_{UNG})$	$lag(X_{UNG})$ represents the lagged close price of UNG
$lag(X_{MPC})$	$lag(X_{MPC})$ represents the lagged close price of MPC

### 3. Results

#### 3.1. Multiple Linear Regression Models' Results

The regression calculation results shown in Table 6 indicated that the daily trading high and low prices of the stock as well as the opening price can be used to predict the closing price of the stock on a particular day. However, the influence of these variables on predicting the stock price of Tesla specifically appears to be relatively limited.

Table 6: The coefficient and P-value of based on the multiple linear regression for Tesla.

Coefficient	P-value
$\beta_0 = 1.458e-01$	P-value = 0.487
$\beta_1 = 1.005e+00$	P-value = <2e-16
$\beta_2 = 6.438e-01$	P-value = <2e-16
$\beta_3 = -6.552e-01$	P-value = <2e-16
$\beta_4 = -1.208e-09$	P-value = 0.171

By calculation, the results of UNG and MPC were also obtained as shown in Table 7.

Table 7: The coefficient and P-value of based on the multiple linear regression for UNG and MPC.

Coefficient	P-value
$\beta_0 = 247.2331$	P-value = <2e-16
$\beta_1 = -18.4217$	P-value = <2e-16
$\beta_2 = 2.5522$	P-value = <2e-16

The multiple linear regression study results indicated that the closing prices of both UNG and MPC can be used to predict the closing price of TESLA on the same day, and the closing price of UNG has an inverse correlation with TESLA and the closing price of MPC has a positive correlation with TESLA.

#### 3.2. Polynomial Regression Models' Results

Through the application of polynomial regression modeling, the high and low prices as well as the opening price, the study demonstrated that all three data can have some prediction for TESLA's close stock price; however, the accuracy of these predictions, particularly when relying on the opening price, is limited, as evidenced by the root mean square error (RMSE) of 4.008556. However, the  $R^2$  values of all three models are close to 1, which indicates that the three variables of Tesla's stock price, the high, the low and the open price can explain the close price of the stock on that day very well. Therefore, it is reasonable to speculate that these three variables, the highest stock price, the lowest



stock price, and the open price combined have an obvious non-linear relationship with the close price of Tesla on the particular day.

### 3.3. Lag Regression Models' Results

By using the formula of lag regression model, this study calculates the coefficients and the p-values of each coefficient shown in Table 8.

Table 8: The coefficient and P-value of based on the lag regression.

Coefficient	P-value
$\beta_0 = 4.263882$	P-value = 0.00623
$\beta_1 = 0.990146$	P-value = $< 2e-16$
$\beta_2 = 0.008364$	P-value = 0.73056
$\beta_3 = -0.227020$	P-value = 0.03229

According to the Lag regression model, the coefficient of MPC lagged close price is 0.008364 and the P-value of it is 0.73056, the coefficient of UNG lagged close price is -0.227020 and it has a P-value 0.03229. Then, it shows there is a relatively significant opposite relationship between UNG lagged close price and TESLA stock price, but MPC lagged price does not explain TESLA stock price very well.

Accurate prediction of TESLA's stock price, specifically the closing price, necessitates the inclusion of daily trading data encompassing the high, low, and open prices of TESLA stock itself, along with relevant data pertaining to Tesla-related stocks to improve prediction accuracy. The research findings suggest that new energy stock prices are minimally affected by crude oil stock price fluctuations, but are more affected by natural gas fluctuations, and that new energy and natural gas stock prices move in opposite directions. Consequently, it can be inferred that the natural gas market holds the potential to influence the new energy market. From long term, the fluctuation of oil company stock price will not affect Tesla stock and will only have a small impact on Tesla stock price in the short term.

## 4. Discussion

Although from an economics perspective, oil stock prices should move in the opposite direction of new energy, in reality, oil stock prices are not good predictors of new energy stock prices, probably because people are investing in new energy stocks for the long term to promote their continued improvement of the global fossil energy shortage. Therefore, the volatility of international crude oil can only bring about short-term shocks in the new energy market [6].

This study ignores the study of other influencing factors, such as the stocks of high-tech companies largely affect Tesla stock. Because Tesla is not an energy-based company in the pure sense, but a technology-based company, the stock market of other related technology-based companies in the market should also be studied to better explain the movement of Tesla's stock price [6]. In other words, studying stocks in the energy sector alone is not enough.

Stock prices are also influenced by many other factors, among them the investment preferences of investors and market news and information [3]. A company's stock price will only rise if investors have high expectations for the company's stock price. In other words, investors believe that the company has great potential for future growth and therefore choose to invest their money in the long term to get high returns in the future, while for the company, financing becomes more available and there is more money to develop itself and the share price will rise in the future. However, in this research, the investor preferences were not considered, which deserves more attention. Additionally,

some more advanced machine learning models such as neural networks can be also considered in further improving the performance of the prediction due to their satisfactory performance in various tasks [7-10].

## 5. Conclusion

In this study, Tesla daily stock trading data is examined to be separate from the impact of MPC, UNG stock on Tesla stock. And the study found that Tesla stock price is affected by the stock price of natural gas companies (UNG) as well as the daily trading of Tesla stock. Based on the results of this study, it is possible to find a general connection and regularity between Tesla's stock price and some economic and financial data. In the future, the results of this study can be combined with the economic environment data to find the effect of the general environment on the stock price if more complex and significant Tesla stock prediction models are to be studied further. In addition, the findings of this study can be used to predict other stocks, not only TESLA, and the process and methodology of this study can be applied to other stock price data if the corresponding data sets are available.

## References

- [1] Zou, C., Zhao, Q., Zhang, G., et al. (2016) *Energy revolution: From a fossil energy era to a new energy era*. *Natural Gas Industry B*, 3(1): 1-11.
- [2] Harding, M. C., Lamarche, C. (2021) *Small steps with big data: using machine learning in energy and environmental economics*. *Annual Review of Resource Economics*, 13: 469-488.
- [3] Strauss, N., Smith, C. H. (2019) *Buying on rumors: how financial news flows affect the share price of Tesla*. *Corporate Communications: An International Journal*.
- [4] Dutta, A. (2017) *Oil price uncertainty and clean energy stock returns: New evidence from crude oil volatility index*. *Journal of Cleaner Production*, 164, 1157-1166.
- [5] Thuan, N. D., et al. (2022) *Forecasting the Opening and Closing Price Trends of Stock Using Hybrid Models and Artificial Intelligence Algorithm*. *Future Data and Security Engineering. Big Data, Security and Privacy, Smart City and Industry 4.0 Applications: 9th International Conference, FDSE 2022, Ho Chi Minh City, Vietnam, November 23–25, 2022, Proceedings*. Singapore: Springer Nature Singapore, 532-546.
- [6] Qu, F., et al. (2021) *Is New Energy Driven by Crude Oil, High-Tech Sector or Low-Carbon Notion? New Evidence from High-Frequency Data*. *Energy*, vol. 230, p. 120770
- [7] Al-Shayea, Q. K. (2011) *Artificial neural networks in medical diagnosis*. *International Journal of Computer Science Issues*, 8(2): 150-154.
- [8] Yu, Q., Yang, Y., Lin, Z., et al. (2020) *Improved denoising autoencoder for maritime image denoising and semantic segmentation of USV*. *China Communications*, 17(3): 46-57.
- [9] Beltratti, A., Margarita, S., Terna, P. (1996) *Neural networks for economic and financial modelling*. London, UK: International Thomson Computer Press.
- [10] Kaastra, I., Boyd, M. (1996) *Designing a neural network for forecasting financial and economic time series*. *Neurocomputing*, 10(3): 215-236.