**A Neural Network Model for Natural Gas Price Prediction**

Jinghua Cao, Jan. 2022

**ABSTRACT**

The model explores using machine learning technic to perform deep learning on natural gas industry trends and consumer demands, and predict natural gas future prices movement. It experiments the possibility to use advanced computer algorithm in structural modeling to predict human behavior, such as the pricing of a cyclical commodity. It will also compare the model result from different machine learning algorithms, along with statistical modeling.

1. **INTRODUCTION**

Natural gas future price has historically been very volatile due to seasonal demand and supply imbalance, the extreme weather conditions, as well as the speculation on the uncertainties of the future demand and consumption needs. However, the future contracts have expiry dates and few speculators are willing to take physical deliveries, so the effect of speculation will cancel out overtime. Economic activities play a key role in determining the price movement of natural gas futures in the mid-term and over decades. For example, due to the shale gas boom, natural gas has been in a downtrend for over a decade although more and more natural gas has been used in electric power generation.

Natural gas is mainly methane, a strong greenhouse gas. It cannot be released into the atmosphere directly due to strict government regulations and industry standards. When the current price level makes natural gas producers uneconomical to transport the natural gas for sale, it has to be burned at well sites because CO2 produces less greenhouse effect than methane. Neither can natural gas be transported overseas easily. There are only limited port and LNG ship capacities to send natural gas over the Pacific and the Atlantic ocean. The current capacity is at maximum 11-13Bcfs per day, versus 100+ Bcf production per day. Global surplus or shortage has limited impact in local market prices.

For the reason, natural gas has been contained in a closed container since extraction from the site, either a transportation pipe, or on site or underground storages, before it is burnt in a residential household or a power plant. It can be accounted for and we can build a decent picture of when and where the natural gas is used over multiple decades. A model can be built to investigate the small imbalance between demand and consumption and predict the price movement in the market.

Machine learning models are able to solve complicated non-linear regression problems with great success. Artificial neural networks(ANNs) has performed remarkably in pattern recognition and data analysis.

* 1. Natural Gas Industry

Natural gas is a naturally occurring hydrocarbon, a class of organic compounds. Raw natural gas from the wells also contain non-energy components that have to be removed in processing plants before the natural gas is marketed and placed into pipelines. In the past, natural gas was a byproduct of oil production. During recent years, while natural gas production from conventional wells has been gradually depleting, natural gas production from unconventional sources, such as shale gas supplies, has increased substantially. As a result, annual U.S. Lower 48 natural gas production increased from 55.1 Bcf/d in Jan 2000 to 99.6 Bcf/d in Dec 2020. Shale gas now accounts for over 70% of Lower 48 natural gas production.

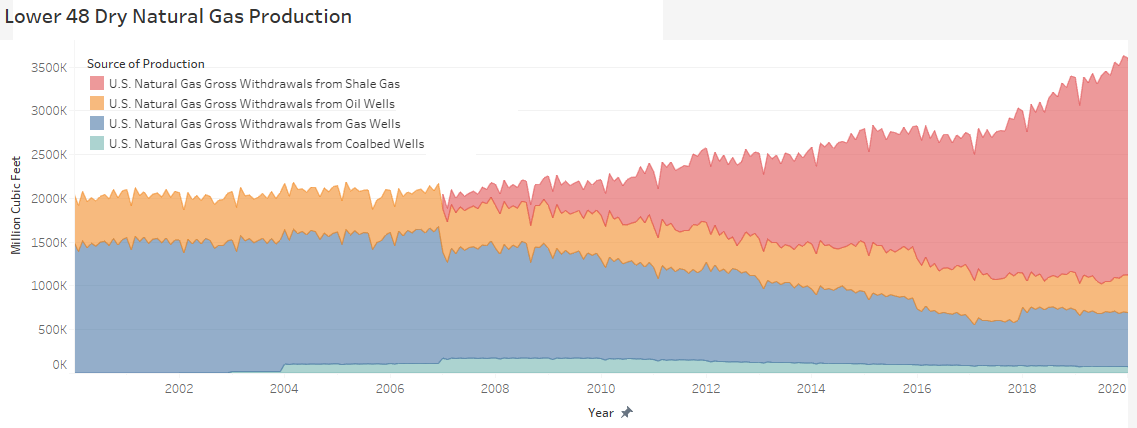


Figure 1 Dry Natural Gas Gross Withdrawals in Lower 48 States by Production Type

The main uses of natural gas are industrial, residential and power generation. Industrial users often use natural gas as a source of heat. Natural gas furnaces ignite and turn off quickly, comparing to coal furnace. Residential households use natural gas for home heating in the winter, as well as furnaces and water heaters. Power plants are the fastest growing users of natural gas because natural gas is more environmental friendly than coal or oil-based plants. In June 2021, 58% of natural gas production was used in power generation. It accounted for 40% of the electricity generated that month.

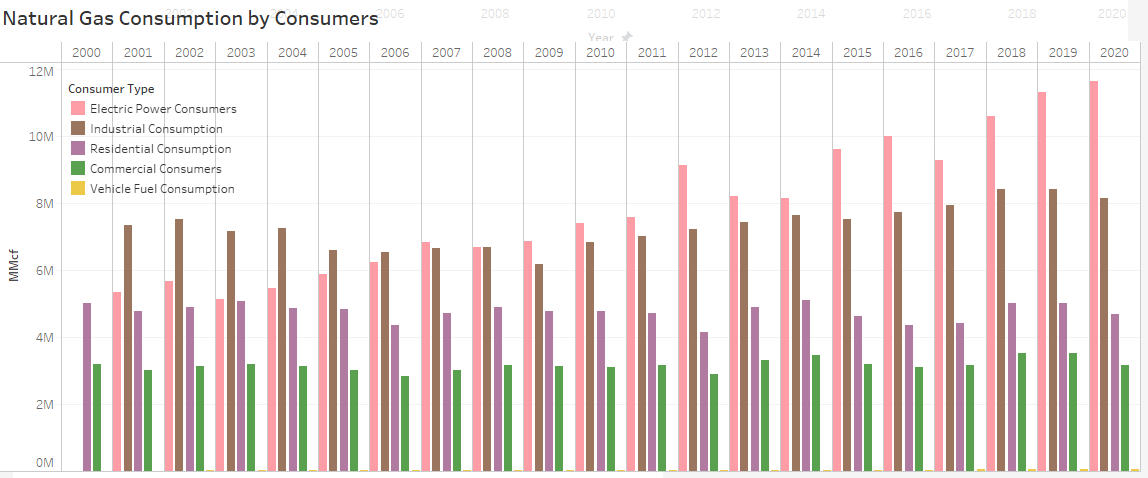


Figure 2 Annual Natural Gas Consumptions in Lower 48 States by Consumer Type

In the most of past 40 years, U.S. imported natural gas by pipeline from Canada to fill the gap between consumption and production. With the growth in domestic shale gas supplies, U.S. has become a net exporter of natural gas since 2017. In Oct 2021, U.S. net exports 10 Bcf/d, mostly due to net LNG exports to European and Asian markets.

* 1. Natural Gas Marketplace

The domestic natural gas marketplace has highly active spot markets and derivative markets. Across continental U.S. and Canada, brokers and others buy and sell natural gas for physical delivery in local delivery points. The benchmark price for North American natural gas is the Henry Hub, located in southern Louisiana. It is interconnected with 13 different interstate pipelines. When quoted by a trader, the price is the difference between the Henry Hub price and that location’s price, called the basis price.

Future contracts are traded in both NYMEX and ICE. NYMEX mainly trades Henry Hub futures. A NYMEX natural gas futures contract requires the seller to deliver and the buyer to take the delivery. Futures trading in NYMEX has grown substantially since its inception in 1990. ICE trades futures on major spot hubs (including Henry Hub) through its U.S. Energy Division (IFED). ICE futures are cash settled based on the monthly price published by NYMEX (Henry Hub) and FERC (other major hub futures). However the open interests in all ICE natural gas future products are often 4 times larger than the open interests in NYMEX.

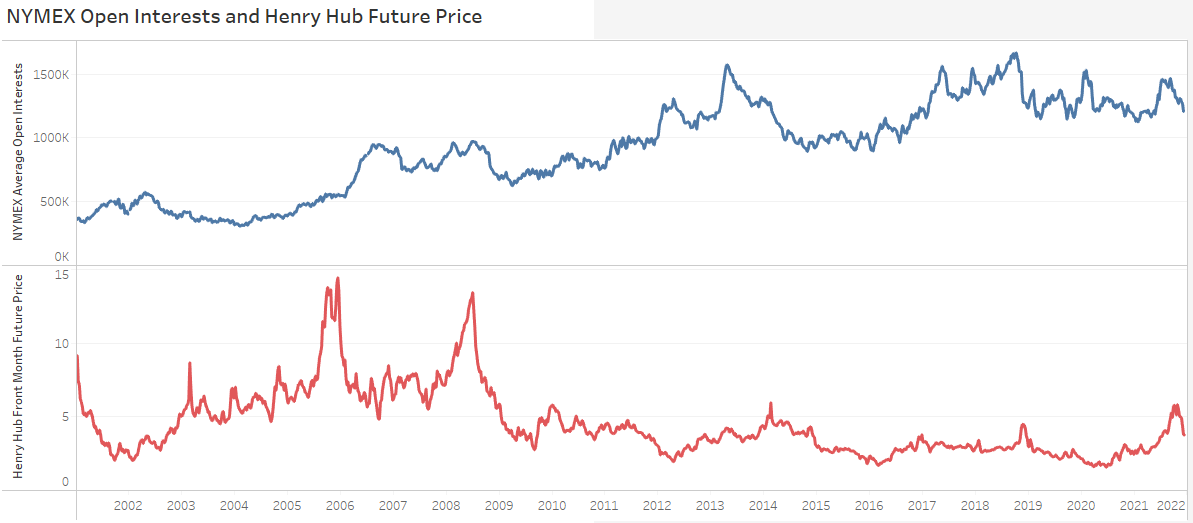


Figure 3 Weekly NYMEX Henry Hub Future Open Interests and Front Month Future Prices

1. **METHODOLOGY**

The presented method is to predict month end henry hub future price by the industry activities and climate change. Specifically, it investigates the production of natural gas, the consumption of natural gas by different consumer types, the import and export during the period, the storage movement, the weather data and the market speculation level to predict month over month price changes. As with other machine learning methodologies, the process consists of three main steps: data collection, data preprocessing and feature extraction, and regression prediction.

* 1. Data Collection

Data were collected through multiple government agencies’ publications. The production breakdowns, the consumption breakdowns, the inventory data, the import and export data are collected through EIA’s historical archives. Heating Degree Days(HDD) information was collected through NOAA’s archives. Historical henry hub future price were using NYMEX Henry Hub Future daily closing prices. The market speculation level was using CFTC’s weekly Commitment of Traders (COT) report. The full data descriptions were included in Appendix I.

* 1. Data Preprocessing
  2. Machine Learning Process

Artificial Neural Networks(ANN) is a basic architecture of a neural network, which consists of layers of neurons connected densely. This type of networks is also known as Multi-layer Perceptrons(MLP), or the “vanilla” neural networks. An MLP neural network is a feed-forward multilayer network architecture that consists of three types of layers, an input layer, multiple hidden layers and an output layer. Except for the input nodes, each neuron in other nodes has its own non-linear activation function. MLP utilizes a supervised learning technique called backpropagation algorithm to train the network.

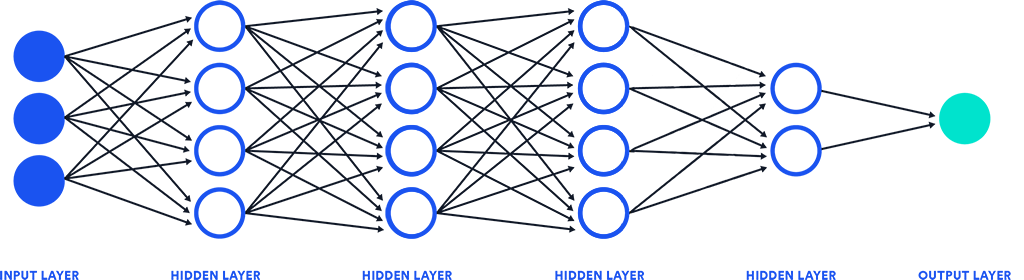


Figure 4 Multi-Layer Perceptron Neural Network Architecture

The input layer has 32 neurons, corresponding to 32 input dimensions. The output layer has one neuron, it predicts the price movement of the month. The learning process takes place in 4 hidden layers. Each neuron in the hidden layer is connected to all the neurons in the previous layer. Each neuron is an individual cell and calculates its own output from all outputs in the previous layer,

, where

Each neuron makes its own decision through an non-linear activation function called Exponential Linear Units(elu). Compared to other activation functions, elu function graduately dims in the network and provide a smooth transition in the regression modeling.

|  |  |  |  |
| --- | --- | --- | --- |
| Activation | *linear* | *relu* | *elu* |
| Function |  |  |  |
| Graph | _images/linear.png |  | https://ml-cheatsheet.readthedocs.io/en/latest/_images/elu.png |
| Pros | * It gives a range of activations | * Avoids and rectifies vanishing gradient problem * Less computationally expensive | * Smooth slowly until its output equal to –α * Can produce negative outputs |
| Cons | * Derivative is constant * Constant gradient descent | * Some neuron dies. * Gradient will be 0 when x<0 * For x>0, it can blow up the activation | * For x>0, it can blow up the activation |

Table 1 Comparison of Activation Functions in Regression Models

Learning is achieved through the backpropagation algorithm. The backpropagation algorithm calculates the cost of the network in each example, and uses gradient descent to minimize the cost function and optimize the network. The neural network uses mean squared error as the cost function for all the hidden layers. The optimization process applies gradient descent to optimize the weights and biases in each neuron. The backpropagation evaluates the derivatives of the cost function from the last layer backwards through the whole network, computing the gradient at teach layer, then perform gradient descent of the whole network to minimizes the error. The learning completed when the loss has been relatively minimized at a low level.

* 1. Training and Validation

The process splits the data into pre-2020 set and post-2020 set distinctively for training and testing purpose. It uses the complete 20 year history from Jan-2000 to Dec-2019 as the training set to train the neural network and use the observations of month over month price changes since Jan-2020 to test the model. The neural network has no information of the upcoming black swan event, a virus was going to swipe through human societies in two months and cause significant psychological shift in the market and complicated economic impacts in the next two years.

1. **RESULTS AND EVALUTION**

The training process runs 1000 cycles (epochs). Usually the neural network is able to minimize the loss in its stochastic gradient descent and finish optimization within the first 20 epochs. The longest observed run took about 40 epochs to finish optimizing the network. Using training data set, we observed the mean squared error(MSE) and correlation coefficient between the output and the observations after each optimization,

|  |  |  |
| --- | --- | --- |
| Epochs | Mean Squared Error | Correlation Coefficient |
| 001 | 0.013 | 32% |
| 003 | 0.012 | 42% |
| 004 | 0.012 | 46% |
| 007 | 0.011 | 42% |
| 008 | 0.011 | 40% |
| 009 | 0.010 | 41% |
| 010 | 0.010 | 43% |
| 011 | 0.009 | 45% |
| 027 | 0.008 | 62% |

Table 2 Model Optimization in Training Data

After the training, the model is used to predict the month end henry hub natural gas future price from the beginning of 2020 to the end of 2021.

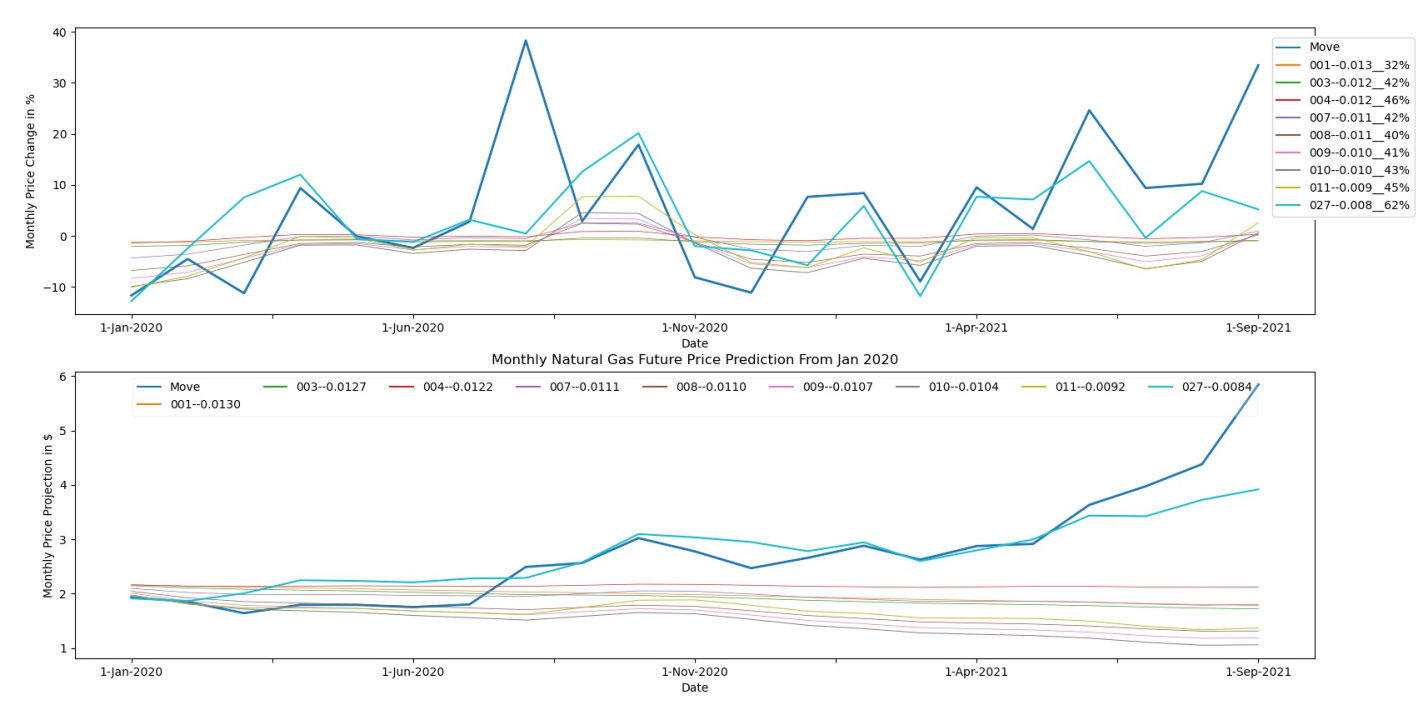


Figure 5 Model Prediction on Front Month Henry Hub Future Price Since 2020

* 1. Result Analysis

The navy line was the observed month end future price in NYMEX. The line marked from 001 to 027 reflected the prediction outcome from different stages in the learning process. 001 curve represents the beginning of the learning process when the neural network was poorly trained. 027 curve marks the completion of the learning process when the network is well optimized and finalized.

The predicted price change curves at the top of Figure 5 demonstrated, the 1st two learning outcomes 001 and 003 have barely discovered the key factors determining the future price. In the third optimization, 004 correctly identified the seasonality in the future price movement that it usually rise before the winter and fall in the spring (it has no knowledge what a spring or winter is). After 004, the learning curve accelerated. The neural network was able to reinforce the knowledge and identified more key factors. The neural network was compiled to use mean squared error (MSE) as the evaluation metric. There is an interesting finding in the learning process. The MSE has been declining from 004 to 009 while the correlation coefficient between observation and output was also declining. Using the curves in testing results as guidance, the algorithm did the right thing. Instead of estimating the direction of each price movement correctly, the algorithm tries to better understand the factors behind the large price movements. After understanding the factors in large price movements, the correlation between observation and output resumes to rise when the optimization goes on. After 011, it took another 16 epochs for the algorithm to find the next gradient descent direction and dive into the local bottom on 027.

The cyan color curve at the top of Figure 5 is the prediction made by finalized and fully trained neural network model. Comparing it to the navy color curve which is actual observation of month over month price movement, the neural network correctly predicted most of large price jumps and dips over the two year period.

The cyan color curve at the bottom of Figure 5 is the cumulative price movement on top of actual price on Dec 2019. It represents the model predicted price level on continuous projections. Comparing it to the navy color curve which is actual observations of the henry hub future prices, the neural network was able to constantly predict reasonable natural gas price all the way to the end of 2021.

* 1. Model Limitations and Discussion

The artificial neural network is part of supervised learning models. It can only predict outcomes based on the information we feed to the model. If the required information is not present in the data, the model will not be able to predict reasonable outcome.

For example, there are a few month the model predict little price change during the month, however, the actual price changed jumped/dropped dramatically. One of the months is August 2020. The neural network, no matter how we train it, what stages of the training it is in, has an inconvincible confidence there will be little price change in the month. However, in reality, the future price jumped more than a third, from 2.4 to 3.2 in a month. There is a story behind it. In August 2020, EIA reports that more than 60% of natural gas production in July was used for power generation. It was a historical record. Natural gas has become the biggest fuel source of power generation, accounts for more than 36% of the electricity generated in July. There is other good news in August too. The expectation of upcoming vaccine increased market optimism that pandemic will end in the end of 2020. And also, oil price recovered from -33 and stabilized in mid 30 level. The market sentiment changed in August without any extraordinary activities we can observe in our data set. However, what neural network read in August was, the consumption in power generation peaked in July at 1300Bcf and dropped 30% to a total of 1000Bcf in August. It is negative information and the price should not move at all.

In another month, Jan 2021 month end, the neural network unanimously predicts the price should drop, while the actual price zigzagged and ended the month with 10% increase. In January 2021, there were sudden stratospheric warmings. However, by the end of January, meteorologists had warned a polar vortex was coming. The actual future price dropped to the low in the mid of January and quickly recovered on renewed forecasts. However, the neural network, read the monthly average HDDs, which has been averaged by the warmer days, did not expect weather pattern change coming and predicted price drop due to the warmer weather, lower residential consumptions and less storage withdrawals.

This neural network model, also constantly underestimate the human greed level and how parabolic a price movement can be. Every training that is able to predict 2005 and 2008 paraboric price movement turned out to be poorly trained model in overall prediction. Every well trained model that is able to predict price movement before and after 2020 most of the time, is very “shy” in predicting 2005 and 2008. While the real movement can be 50% change month over month, the model only predicts 25% or 30%. So does the model predict the future price in Sept 2021. The model agrees the price ought to rise, but not like that. In this data set, we lack a better market sentiment measure that can reflect once/twice-in-a-decade market frenzy. Or we can just respect an artificial neural network’s view on the craziness of actual neural networks.

1. **MODELS COMPARISON AND DISCUSSION**

We compared the outcome from neural network models with other popular machine learning algorithm, as well as the traditional linear regression model. The result is presented in the figure below,

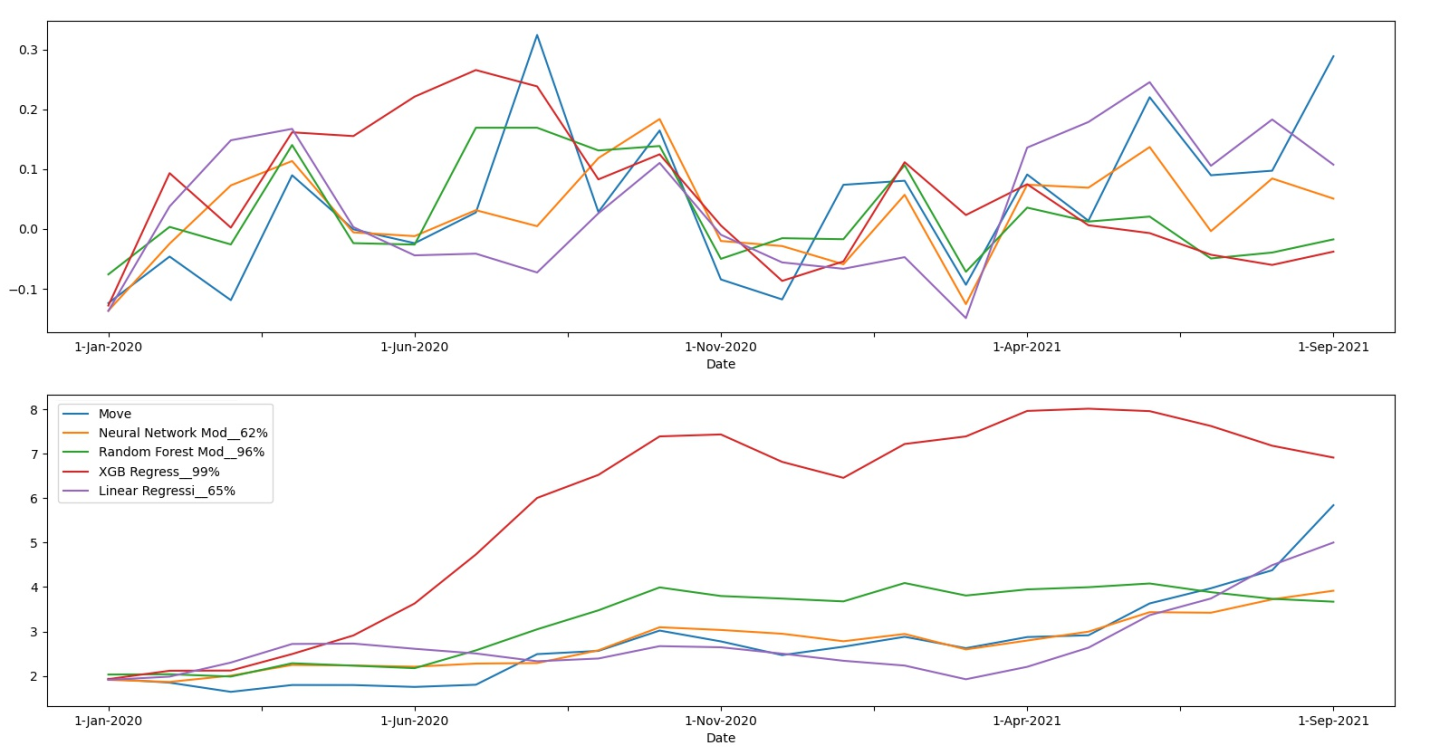


Figure 6 Comparison of Machine Learning and Statistical Models

1. **CONCLUSION**

**APPENDIX I. DATA DESCRIPTION**

|  |  |  |
| --- | --- | --- |
| **Data Field** | **Source** | **Description** |
| Last Future | NYMEX | Last Period Front Month Future Price |
| Dry Production | EIA |  |
| Residential Consumption | EIA |  |
| Commercial Consumption | EIA |  |
| Industrial Consumption | EIA |  |
| Electric Power Consumption | EIA |  |
| Δ(Dry Production) |  | Dry production change over the last period |
| Δ(Residential Consumption) |  | Residential consumption changes over the last period |
| Δ(Commercial Consumption) |  | The change of natural gas deliveries to commerical consumers over the last period |
| Δ(Industrial Consumption) |  | Industrial consumption changes over the last period |
| Δ(Electric Power Consumption) |  | The change of natural gas deliveries to electric power consumers over the last period |
| Pipeline Imports | EIA |  |
| Pipeline Exports | EIA |  |
| Liquefied Exports | EIA |  |
| Working Storage | EIA |  |
| Net Withdraws | EIA |  |
| Δ(Pipeline Imports) |  |  |
| Δ(Pipeline Exports) |  |  |
| Δ(Liquefied Exports) |  |  |
| Δ(Working Storage) |  |  |
| Δ(Net Withdraws) |  |  |
| Non-Comm Net Stded | CFTC |  |
| Δ(Non-Comm Net) |  |  |
| HDD CONUS | NOAA | Average Heating Degree Days(HDD) in lower 48 states |
| Δ(HDD CONUS) |  |  |
| Month |  | Calendar reporting month of the year corresponding to above economic data period |

**REFERENCES**

Haykin, S. O. (2011). Neural networks and learning machines (3rd ed.). New York: Prentice Hall.