nbvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvcbcbcbcbcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

Mastura Mir, Jas Sur, Tracy Wells, Vandana Yarala, Patrick Cardoso

mirm1@gator.uhd.edu; surj1@gator.uhd.edu; wellst19@gator.uhd.edu; yaralav1@gator.uhd.edu; cardozop5@gator.uhd.edu

Abstract

Henry Hub, based in Louisiana, is a key trading location because it is the benchmark used to set prices for all the other trading locations across the country.  As a result, daily Henry Hub prices are carefully tracked and archived by the Energy Information Administration (EIA), a branch of the United States government. Daily natural gas price and date (for seasonality) are the only variables we expect to use in this work.    
Projecting the near-term price movement of natural gas is useful for both speculative traders and physical traders who manage the actual commodity and its transport across the country. The goal of any trader or risk manager is to buy gas at low prices and sell at high prices as the market moves up or down in response supply and demand at different locations.  In this project we intend to answer whether time series modeling offers useful predictive capabilities related to Henry Hub price movements. If the model is successful, it will help traders, producers, and consumers to be profitable by predicting potential price movement ahead of when it occurs. We analyze various models to use an integrated Autoregressive / Moving Average model type (ARIMA). We compare the Ljung Box Q-statistics p-value for residuals to check for null hypothesis of residuals being white noise. We also check Auto Correlation Function and Partial Auto Correlation Function that may be indicative of model selection. In the end we compare various models based on SBC, AIC and model parameters to select a model and predict natural gas prices for near future

time series anslysis of natural gas spot prices

MS(Data Analysis)

University of Houston Downtown

Table of Contents

[Time Series analysis of Henry Hub Natural Gas Spot Price 2](#_Toc87776536)

[Apply Difference (1) 3](#_Toc87776537)

[ARIMA model comparison 4](#_Toc87776538)

[SBC Ranking 4](#_Toc87776539)

[AR(1) 4](#_Toc87776540)

[ARI(1) 6](#_Toc87776541)

[IMA(1,1) 7](#_Toc87776542)

[ARIMA(1,1,1) 8](#_Toc87776543)

[Other potential models 10](#_Toc87776544)

[AR(2,1) 10](#_Toc87776545)

[ARIA(3,1,2) 12](#_Toc87776546)

[Outliers check 14](#_Toc87776547)

[Comparison of ARIMA(1,1,1), ARMA(2,1), ARIMA(3,1,2) 15](#_Toc87776548)

[Final model Selection 16](#_Toc87776549)

# Time Series analysis of Henry Hub Natural Gas Spot Price

The data appears to not follow any specific pattern. It has no constant mean, and by looking at ACF, the lag doesn’t decay quickly. So we decided to use a Difference to see what it entails

Chart, scatter chart

Description automatically generated

Table

Description automatically generated

## Apply Difference (1)

Chart, scatter chart

Description automatically generated

Table

Description automatically generated

The residuals have constant mean and variance is normally distributed. ACF has a spike at lag 1 and then it appears to be following a sinusoidal pattern but looking at the p-values; some of the values suggests that out null hypothesis that residuals is white noise is rejected, so this model isn’t a valid model.

We decided to do an Arima Group model using a p,q,d or 2,2,1

## ARIMA model comparison

Let’s look at various suggested models based on SBC and AIC ranking

### SBC Ranking

Graphical user interface, application, table

Description automatically generated

#### AR(1)

ARIMA model group comparison based on the SBC Rank suggests that AR(1) has a rank of 1, so let’s look at its residuals and ACF to verify if this a valid model.

Chart, scatter chart

Description automatically generated

The residuals appear to have constant mean of zero and the variance appears to be normally distributed.

Table

Description automatically generated

Table

Description automatically generated

The ACF has a spike for lag 1 and then it follows a sinusoidal pattern. However, the pvalues rejects the null hypothesis that the residuals is white noise, so this is not a valid pattern.

#### ARI(1)

Chart, scatter chart

Description automatically generated

The residuals have a constant mean, and variance is normally distributed and doesn’t follow any specific pattern.

Table

Description automatically generated

ACF has a spike at lag1, and the follows a sinusoidal pattern. However the pvalues reject the null hypothesis that the residuals is white noise, so this model isn’t valid.

#### IMA(1,1)

Chart, scatter chart, box and whisker chart

Description automatically generated

The residuals have a constant mean of zero and the variance appears to be normally distributed and doesn’t follow any specific pattern.

Table

Description automatically generated

The ACF values has a spike at lag 1 and none afterwards. It also follows a sinusoidal pattern. However, the pvalues rejects the null hypothesis that the residuals is white noise. So, we reject this model as well.

#### ARIMA(1,1,1)

Chart, scatter chart

Description automatically generated

The residuals have a constant mean of zero and the variance appears to be normally distributed and doesn’t follow any specific pattern.

Table

Description automatically generated

Table

Description automatically generated

The ACF plot has a spike at lag1, and none afterwards. It also follows a sinusoidal pattern. The Q statistics pvalues accepts the null hypothesis that the residuals are white noise. So, this is a valid model.

ACF plot also suggests that it has no seasonality, so don’t need to apply a seasonality check.

Graphical user interface

Description automatically generated

### Other potential models

Based on the ACF plot and testing the Q-Statistics rvalues that accepts the null hypothesis, only the following models qualify. These models have valid ACF plot, with a spike at lag1 and follows a sinusoidal pattern afterwards. The rvalues accepts the null hypothesis for residuals being white noise.

#### AR(2,1)

Chart, scatter chart

Description automatically generated

Table

Description automatically generated

A picture containing table

Description automatically generated

#### ARIA(3,1,2)

Chart, scatter chart

Description automatically generated

Table

Description automatically generated

A picture containing graphical user interface

Description automatically generated

### Outliers check

## Comparison of ARIMA(1,1,1), ARMA(2,1), ARIMA(3,1,2)

Graphical user interface, application, table, Excel

Description automatically generated

ARIMA(1,1,1) has AIC rank of 6, and SBC rank of 5, while ARIMA(3,1,2) has a AIC rank of 5 and SBC Rank of 16, so it’s clear that ARIMA(1,1,1) is a better model in comparison to ARIMA(3,1,2). ARIMA(1,1,1) has a lower p,q,d values so it’s parsimonious as well.

Comparing ARIMA(1,1,1) with ARMA(2,1). The former has AIC rank of 6, and SBC rank of 5, while the latter has a AIC rank of 1 and SBC rank of 6. So, we should check the model data and forecast to check which one may be a better model.

|  |  |
| --- | --- |
| ARMA(2,1) | ARIMA(1,1,1) |
| A picture containing table  Description automatically generated | Graphical user interface  Description automatically generated |

ARMA(2,1) have a smaller confidence interval in comparison to ARIMA(1,1,1).

ARMA(2,1) has a valid AR2, MA1, and intercept value (p-value < 0.005), while AR1 value has a p-value of 0.3095.

ARIMA(1,1,1) has a valid AR(1) and MA(1) because of pvalue (0.0001 < 0.005) but it’s intercept isn’t valid because of pvalue(0.8789 >0.005).

## Final model Selection