Natural Gas Price Analysis

JONGBAE YOON

Applied Mathematics
UNIVERSITY OF COLORADO BOULDER

Introduction

Natural gas has various advantages over other types of fossil fuels as an energy source, and numerous studies have been conducted to understand not only its supply and demand, but also its price movements. The aim of this study is to comprehend the key predictors of natural gas prices and build a forecasting model using both frequentist and Bayesian methods. To identify the crucial variables, both Vector Autoregression (VAR) and Bayesian VAR (BVAR) methods are utilized. For the forecasting task, Autoregressive Integrated Moving Average (ARIMA) and Bayesian Structural Time Series (BSTS) methods are employed, and the results of each model will be compared using evaluation criteria.

Dataset

The supply and demand dynamics are the fundamental drivers of the price of natural gas. The Henry Hub natural gas price serves as the response variable, with 11 different datasets of supply and demand factors selected as predictors. The dataset covers monthly data from January 1997 to January 2023. Natural log on the dataset to stabilize the time series datasets.

Category		Variable	Unit
Response		Henry Hub Natural Gas spot Price	USD / MMbtu
Supply	Production	Natural Gas Marketed Production	MCF
		Natural Gas Rotary Rigs in Operation	EA
	Storage	Total Natural Gas Underground Storage Vol.	MCF
	Imports & Exports	Natural Gas Imports and exports	MCF
Demand	Weather	Degree days deviation	EA
	Economic growth	Monthly Real GDP Index	USD
	Other fuels	Cushing, OK WTI Spot Price	USD / bbl
		New York Harbor 2 Heating Oil Spot Price	USD / gal
	Consumption	Natural Gas Total Consumption	MCF

Table 1. Dataset

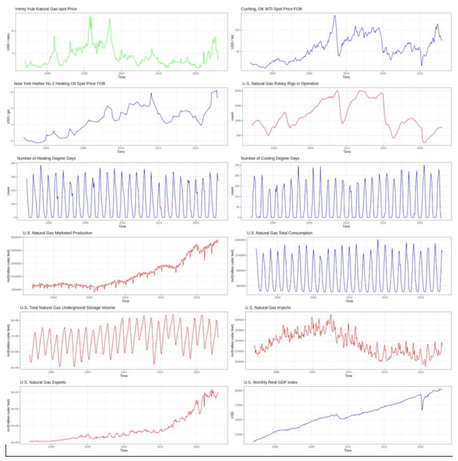


Figure 6. Dataset plot

Methodology

1. Correlation between variables

To find the correlation between variables, VAR and Bayesian VAR (BVAR) are applied, since VAR method is widely utilized to analyze the relationship between multiple time series variables.

$$y_t = c + A_1 y_{t-1} + A_1 y_{t-1} + \ldots + A_p y_{t-p} + e_t$$

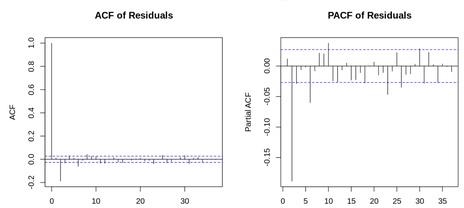
Use Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to find the lag 'p'. BIC = 2 is selected over AIC = 5.

2. Forecasting

For the second goal of building a forecasting model for the future movement of the natural gas price, traditional ARIMA and BSTS models are utilized. ARIMA is one of the most popular time series modeling techniques for the forecasting future values based on past observations.

$$\phi(B)(1-B)^d X_t = \theta(B) Z_t, Z_t \sim WN(0, \sigma^2)$$

ARIMA(1,1,0) is selected; autoregressive term (p =1) and the first order differencing (d=1)



BSTS uses Bayesian methods to estimate the parameters of a time series model and make predictions. In a BSTS model, the time series is decomposed into multiple components, including a local level, local trend, seasonality, and regression component

$$Y_t = \mu_t + X_t + S_t + e_t, e_t \sim N(0, \sigma_e^2)$$

$$\mu_{t+1} = \mu_t + \nu_t, \nu_t \sim N(0, \sigma_\nu^2)$$

The uninformative prior is considered. The model iterates 10,000 times and 10% of burn-in period is applied.

3. Evaluation Criteria

The performance of each model will be evaluated using various evaluation criteria MAE, MSE, RMSE, and MAPE.

$$MAE = \frac{1}{N} \times \sum_{t=1}^{N} |\widetilde{y}_t - y_t|.$$

$$MSE = \frac{1}{N} \times \sum_{t=1}^{N} (\widetilde{y}_t - y_t)^2.$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \times \sum_{t=1}^{N} (\widetilde{y}_t - y_t)^2}.$$

$$MAPE = \frac{1}{N} \times \sum_{t=1}^{N} \left| \frac{\widetilde{y}_t - y_t}{y_t} \right|.$$

Result

1. Correlation

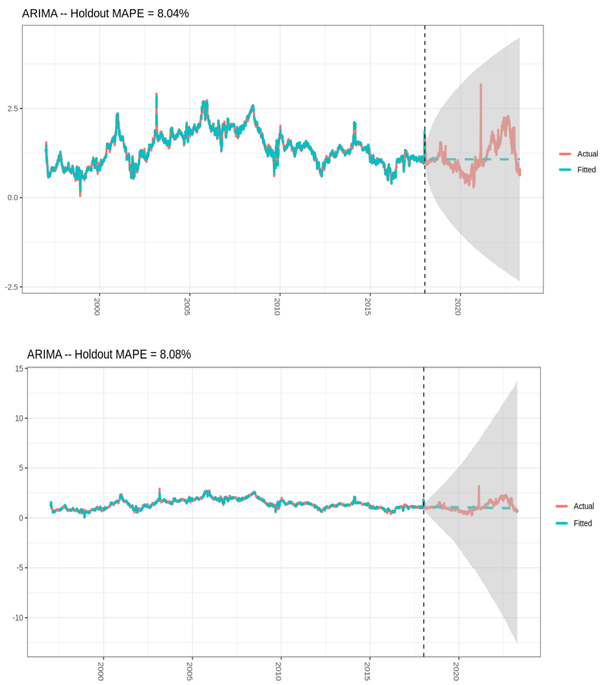
(VAR) The previous natural gas price, heating oil price, monthly gas export, degree days, and the heating oil price of 2 previous time stamp have the significant correlation with the current natural gas price .

(BVAR) The previous natural gas price, WTI crude oil price, heating oil price, natural gas production, GDP have the significant correlation with the current natural gas price .

	VAR	BVAR
MAE	0.34	0.36
MSE	0.19	0.19
RMSE	0.44	0.44
MAPE	0.30	0.39

	ARIMA	BSTS
MAE	0.09	0.09
MSE	0.04	0.04
RMSE	0.20	0.20
MAPE	0.08	0.08

2. Forecasting



Conclusion

The VAR analysis revealed that five factors significantly impact current natural gas prices, including the previous natural gas and heating oil prices, exports, degree days, and heating oil prices two time-stamps before. The previous heating oil price and natural gas price had the strongest correlation. The BVAR analysis returned slightly different results. The five most crucial predictors are the previous natural gas price, WTI crude oil price, heating oil price, natural gas production, and GDP. The VAR analysis had smaller evaluation criteria values than BVAR, indicating that the VAR results were more reliable.

Overall, the evaluation criteria values were acceptable when compared to similar studies. For both tasks, frequentist methods performed better than Bayesian methods. This could be because there was no prior distribution available for the natural gas price and other predictors used in this study, and all variables were assumed to have an uninformative prior. One advantage of Bayesian methods is having prior knowledge, but since no such knowledge was available in this case, the Bayesian method did not perform well.

Limitations, Future works

1. Limitations

- Monthly Dataset
- Missing values
- No Prior information

2. Future works

- Apply Advanced Time Series model
- Apply Artificial Neural Network model
- Apply Recurrent Neural Network (LSTM) model