The Price of Natural Gas:

Understanding Price Movements and Implications for Prediction

Angela Zhang, Martin Zagari, and Omar Elfeky

University of San Diego

Shiley-Marcos School of Engineering

# Abstract

Natural gas is a crucial fuel in modern industrial society, contributing 23% of global primary energy demand, alongside oil, coal, and nuclear fuels (EIA, 2021). According to the US Energy Association, natural gas fuels 40% of US electricity production and that figure continues to increase (EIA, 2021). Nearly half of US households now rely on natural gas for heating (Roberts, 2018). Natural gas is also being increasingly viewed as a “bridging fuel” for climate change: natural gas releases 48% less CO2 than coal and 34% less than oil having equivalent heat content (AEI, 2021). Increasing reliance on natural gas makes consumers, industry, and the economy more vulnerable to one of the main weaknesses in converting energy consumption to natural gas: the higher challenges in transporting, storing, and distributing natural gas to the end user. Natural gas is therefore more subject to local supply and demand factors and market inefficiencies, resulting in more price volatility than oil (Long et al, 2003). For this reason, it is important to study the types of cycles and patterns that characterize the price stability of natural gas if it is to become more important in the transition to a net zero economy. Important research questions include: 1) the importance of seasonality vs market factors in gas price, 2) relationships between gas prices and other energy commodities such as oil and coal, 3) volatility trends in gas market prices.

*Keywords:* commodities, electricity, industry. economy

**Table of Contents**

[Abstract 2](#_Toc75716696)

[Problem Statement 4](#_Toc75716699)

[Natural Gas History 5](#_Toc75716699)

[Data Selection 7](#_Toc75716699)

[Exploratory Data Analysis 7](#_Toc75716700)

[Random Walk Testing 8](#_Toc75716701)

[Model Strategies 9](#_Toc75716710)

Building the baseline ARIMA model [10](#_Toc75716727)

Seasonal ARIMA modeling [11](#_Toc75716727)

[Conclusion 16](#_Toc75716729)

[References 17](#_Toc75716730)

[Appendix A 18](#_Toc75716731)

List of Figures

[Figure 1. Differnced Data Showing Sawtooth Pattern 9](#_Toc75718924)

[Figure 2. Arima(2.1.2) 11](#_Toc75718924)

[Figure 3. Arima(1.1.1) 12](#_Toc75718924)

[Figure 4. Decomposition of Multiplicative Time Series 13](#_Toc75718924)

[Figure 5. Series (diff(diff(monthly)) 15](#_Toc75718924)

[Figure 6. 60 Month Seasonal Forecast of Nat Gas 15](#_Toc75718924)

# Problem Statement

Natural gas appears to be an increasingly important fuel for electricity generation, heating, public and industrial transportation, and is even being proposed as a “bridging fuel” to reduce CO2 emissions. Even after a substantial phase-out of coal for electricity generation in dozens of countries, coal continues to generate nearly 20% of supply of global electricity, suggesting further pressure on natural gas supplies, infrastructure, and price (EIA). Despite rapid adoption of battery power for transport, large scale battery storage of electricity remains prohibitively expensive. And as renewables account for an increasing share of electricity generation, the intermittency of sun and wind put increasing pressure to match supply and demand for fossil fuels. Except in the unlikely case that nuclear power undergoes a renaissance in approval and construction, natural gas will likely be the primary fuel used to expand baseload electricity production.

Modern life and habitation of cold climates relies on the electricity and heating of natural gas, and it is imperative for the oil and gas industry to predict the prices to invest in production to meet demand. Equally critical for high energy users and manufacturers to be able to predict the cost of one of their major inputs. As the recent Experience in Europe has demonstrated, more renewable electricity reduces the use of fossil fuels but does not eliminate it and without careful planning can increase fossil fuel usage. Germany, one of the highest green energy spenders, installed ~54 gigawatts of solar panels enough to technically provide 25% of its electricity needs but only generates 10% (Hancock, E. 2021).

**Natural Gas Price History**

Natural gas prices have exhibited a much different profile than that for other energy sources, particularly oil. At the highest level, natural gas prices exhibited a general upward trend from approximately 1997 through the financial crisis of 2007-2008, and then a general downward trend from 2009 until the present day. Of course, there is substantial volatility in any given week or day. This is different from the generally upward trend in oil prices, meaning that the overall demand for energy and economic growth is likely not the major driver of gas price as it is for oil price. Although the overall natural gas price trend for nearly 25 years has been essentially flat, substantial price movement suggests complex market forces and dependencies that bear exploration.

One important difference between oil and gas is that natural gas is a by-product of oil production, meaning that the demand for and production of oil may very well have a knock-on effect on gas prices. As a result, we will explore the relationship between oil and gas prices, with gas prices potentially being a lagging variable for oil demand and/or production. In particular, the shale oil boom in Texas may have had an outsized impact on the price of natural gas, supressing it for much with last decade with a recent resurgence in prices perhaps being due to a collapse in the shale oil boom.

Another important difference between oil and gas is the relative ease of transport and storage. Despite considerable infrastructure for gas transport by pipeline, there are still many fewer options for natural gas transport and storage than for oil. (But fortunately, natural gas is much more storable than electricity!). Regrettably, lack of infrastructure often means it is cheaper for energy producers to waste natural gas by-products through venting or “flaring” rather than risk oil processing disruptions or pay for liquification, storage, and transport (Hampton, Oct. 11, 2021). For this reason, exogenous shocks such as weather impact gas markets differently than they do for oil markets, and natural gas prices are, in fact, more volatile than for oil (but less volatile than electricity). Evidence of this is seen in headlines speculate on gas price spikes based on overly mild vs severe cold weather, with weather-related price movements being much more common for gas than for oil. (Long, 2003) For this reason, we will also explore the importance of seasonality in natural gas prices that may be obscured by trend and other shocks.

Finally, we still see “shocks” in gas price corresponding with important world and economic events. One gas price spike appears to occur around September 11, 2001. A series of price spikes occurred before and around the financial crisis, perhaps corresponding first to economic growth, followed by economic uncertainty and the collapse of oil prices. Our models will therefore have to account for a high level of exogenous noise in energy markets, as most economic models do.

To summarize, natural gas prices appear to have followed an extremely long cycle upward and downward trend, with frequent macroeconomic and political shocks along this journey, perhaps with underlying seasonality in price. Our analysis will take these domain factors into account.

# Data Selection

# Natural gas price data are widely available from public and commercial sources. One of the first questions for this study was deciding which price data to use and deciding on the density of data observations to model. As is the case for most commodity financial data, data are available for opening, high/low/intraday, and closing prices for every business day, as well as futures and other derivative contracts. Since the purpose of our model is long term trends and patterns, we decided to use less granular monthly data rather than daily data. We also decided that weekly data would not offer enough additional advantage over monthly data since seasonality of price data is most easily understood in terms of a monthly price. Data were downloaded from Kaggle, and the source of the Kaggle data is the continuously updated U.S. Energy Information Administration EIA Data.

# Data Exploration, Pre-processing, and Splitting

Following the importation of necessary packages and the raw datasets, the data was prepared for further evaluation. The initial phase of pre-processing began with inspection of the data and calculation of statistical measures. Other properties such as head and dimension were produced to provide additional information. Statistical measures such as median and mean are useful in outlier detection. However, outliers were not removed from the dataset due to having a potential statistical significance on forecasting results and seasonal behavior. Missing values were collected and handled appropriately. The dataset contained only one missing value which was removed completely. Each data frame was converted to a time series object using the date column. Finally, the resulting time series objects were plotted to visualize both the daily and monthly behavior of natural gas prices.

**Random Walk Testing**

One of the first consideration for pricing data, especially for securities or commodities data, is that the data are a pure random walk, which would make detailed prediction models an impossible task. We wanted to explore this possibility at a high level first before making additional decisions regarding additional analyses.

The ACF plot for the monthly data showed a linearly-decreasing lag series that were large in magnitude and remained significant for many lags, suggesting non-stationarity as well as portending a potential random walk process. We then coerced the data to stationarity using the first difference. After differencing, the ACF showed nearly no lag correlation for any lag, again suggesting a potential for a random walk process.

We used Dickey-Fuller, Augmented Dickey-Fuller, and Phillips-Perron tests for the stationarity of the data and the differenced data. All tests failed to reject the null hypothesis of non-stationarity for the raw data but were significant (rejected the null hypothesis of non-stationarity) for the differenced data. Again, these results in aggregate could not rule out a possible random walk.

Interestingly, the plot of the stationary, differenced data was not entirely consistent with a random process. Looking at this plot (Figure 1) showed some structure and pattern in the data. For example, there appeared to be periods of progressively increasing differenced values, followed by a rapid decline in the values, forming an almost “sawtooth” pattern to these data. This was especially apparent early in the series. For this reason, we were cautiously optimistic that further analysis might reveal some learnable structure to the data such as long-cycle or long-memory processes.

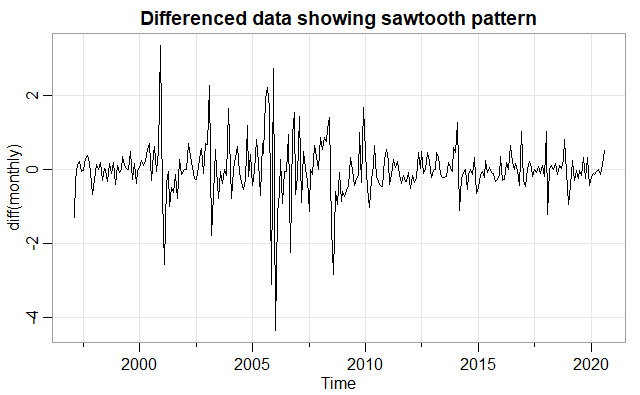


Figure 1

To summarize our thinking, for further analyses and tests, we essentially adopted the stance that the null-hypothesis is that natural gas price data is a random walk. We performed further testing to try to disprove this null hypothesis, especially looking for oscillation, seasonality, or other internal dependencies in the data.

# Model Strategies

An exhaustive review of the literature for commodity pricing is beyond the scope of this research project. There have been numerous publications looking as natural gas pricing, and the content of those publications has become increasingly sophisticated as modelling techniques and computing power have increased. For example, recent publications increasingly apply machine learning techniques such as random forests and neural networks to time series data, presumably based on the rising popularity and high performance of these methods in discerning deep patterns from extensive external data that are increasingly available. We chose to concentrate more on traditional time series approaches, particularly given the transparency of these methods compared with “black box” algorithms.

As noted above, two publications merited our attention based on their recency and focus on time series modelling. Siddique at al compared a baseline ARIMA(2,1,2) model to an Autoregressive model augmented by Neural Network they dubbed “ARNN”. They found that the ARNN model performed 30% better for short term forecasts. For us, the encouraging finding was that an ARIMA model might be useful for our data, even though we were looking for longer term patterns.

A second publication in August 2021 by Berrisch, et al specially addressed making day-ahead and month-ahead forecasts of natural gas price, the month-ahead objective being in line with our data, albeit shorter term than our objective. They were able to incorporate exogenous factors and internal data characteristics (“heavy tails” assumption) to build a state-space model that was 13% (day-ahead) and 9% (month ahead) improved vs the baseline ARIMA model. Although limited by day ahead and moth ahead focus, but that is also where we chose to start since they also started with an ARIMA model by Siddique.

**Building the baseline ARIMA model**

Fitting the ARIMA(2,1,2) model to the long-term natural gas data, we found that all parameters were statistically significant. By trial and error, we tried fitting a variety of models with lower and higher orders. ARIMA(1,0,0) and ARIMA(0,0,1) did not fit the data, but ARIMA(1,1,1) actually appeared to fit the data as well as the ARIMA(2,1,2) with comparable AIC and BIC values, and better Ljung-Box scores. Other models testing having higher orders did not outperform the ARIMA(1,1,1), which became our benchmark model for further analysis.

**Seasonal ARIMA Modeling**

Once we fit the baseline ARIMA model, we turned to the question of seasonal adjustments. Natural gas is well known to exhibit seasonal price movements corresponding to warmer and

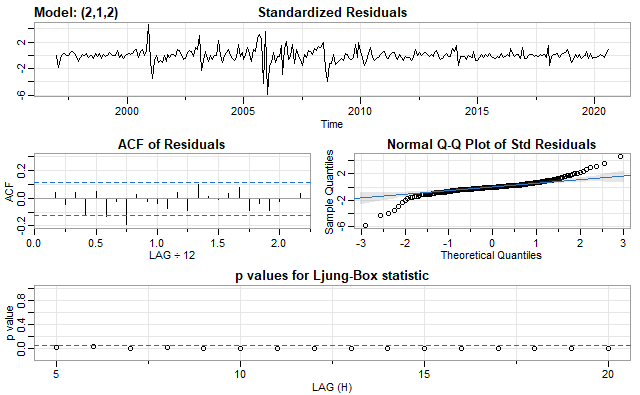


Figure 2

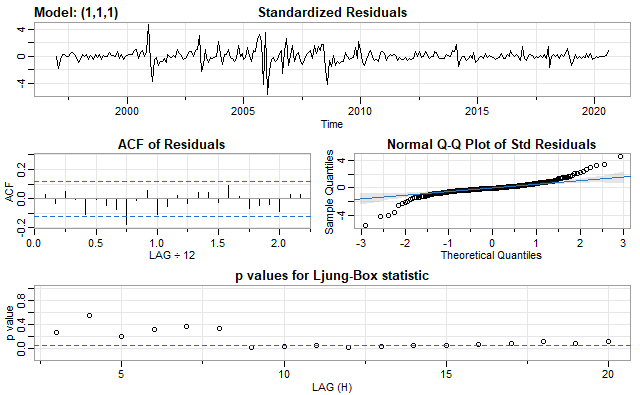


Figure 3

# cooler weather, and these may even be apparent as small yearly-recurring spikes in the raw data. As an initial exploration of seasonality, we ran a decomposition algorithm. Multiplicative decomposition (FIGURE X) was chosen over additive decomposition (results not shown) since the residual errors were much smaller and the seasonal effects much larger. This analysis suggested a small but discernable seasonal pattern was worth making adjustment for in the final ARIMA model.

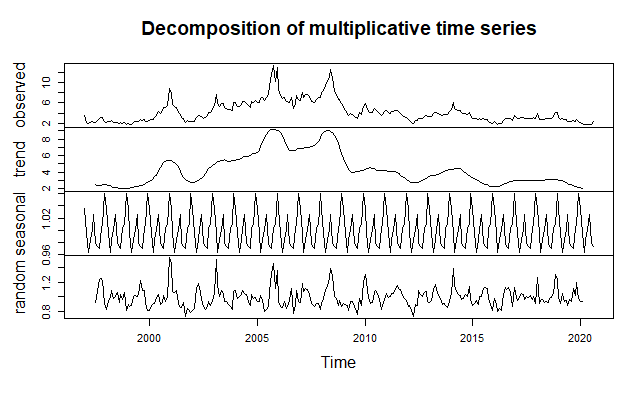


Figure 4

Starting with correlation analysis, we plotted the ACF and PACF of the double differenced monthly natural gas data, expecting to see some evidence of seasonal correlation. Figure 5 shows that there was in fact, strong autocorrelation at 12 months, with the ACF cutting off at lag = 12 months, and the PACF tailing off through lags 12, 24, 36, and 48 months. Together these plots implied an SMA(1) model, i.e. P=0 and Q=1. A periodogram (not shown) also confirmed periodicity in the model, with peaks at 0.04 and 0.08 frequencies, corresponding to 12 and 24 month lags in autocorrelation of the monthly data. By process of trial and error, we learned that the seasonal part of the ARIMA model could be extended to a higher order SARIMA(1,1,1),(1,1,2)[12] model with decent looking residuals.

Finally, we forecast for 60 months to see what the impact of the seasonal adjustment is on the model. As expected, we note that that the seasonal component is small but not negligible when making forecasts (FIGURE X). The seasonal excursions for prices appear to be on the order of approximately $1 due to the seasonal component of the price. Of course, the forecast is not entirely satisfactory since natural gas prices should not decrease below 0 for any appreciable length of time, but the purpose of this research was to demonstrate that natural gas prices are not, in fact, a pure random walk and that seasonal and ARMA characteristics do appear to influence the time series, likely due to recurrent exogenous factors that have been documented and studied by others as noted in the introduction.

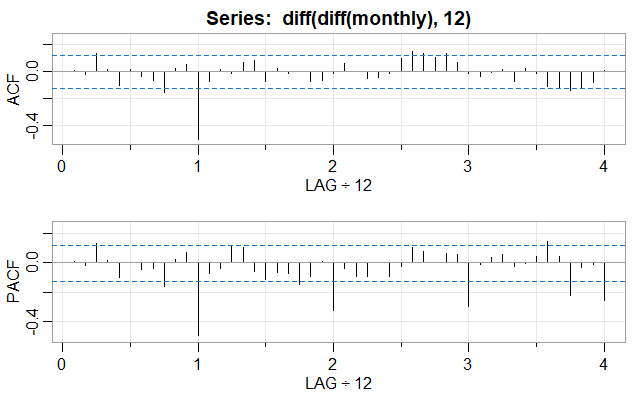


Figure 5

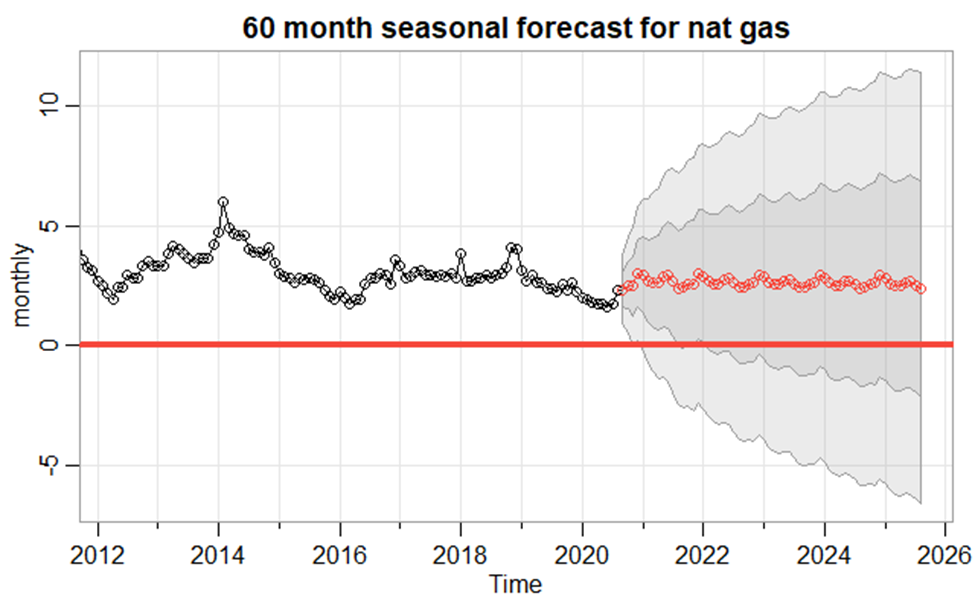


Figure 6

# Conclusion

We set out to analyse long-term patterns in natural gas prices, particularly in light of the increasing importance of natural gas to the global economy. Most interest in commodity pricing in the literature is short term forecasting for trading and hedging, but we set out to investigate if there may be broader patterns in prices that might be useful for policy development, regulatory considerations, and future research in the area. We set out with a null hypothesis that natural gas prices follow a random walk, in keeping with securities price patterns in general, but also knowing that there are seasonal and other considerations for commodities that aren’t factors for pure financial assets. We thought this approach was the most rigorous approach given the high degree of price excursions in gas pricing of 25+ years.

Our research was greatly aided by studies that were recently conducted asking similar questions about daily and monthly price prediction. In particular, the 2019 article by Siddique, et al compared advance neural network modeling with traditional ARIMA techniques, with the result that their Autoregressive Neural Network (ARNN) model was 30% more accurate in predicting gas prices than the baseline ARIMA(2,1,2) model. This gave us the starting point for our own ARIMA modeling of natural gas prices and allowed us to build on that work.

Our results suggest that beyond short-term forecasting, natural gas prices analyzed monthly also follow an ARMA process with seasonal influences. Unlike Siddique et al, we found that the best fit for the data, considering complexity and fit, was an ARIMA(1,1,1) model. Indeed, the nearly identical AIC and BIC scores suggest that the simpler process may be more generalizable as the base model for long term natural gas pricing, and perhaps a more stable platform for model additions.

In terms of model additions, we investigated whether there is an important seasonal component that might be modelled, and all our analyses suggested that there was. An initial exploration using decomposition noted a small 12-month seasonal component. The periodogram also reinforced low frequency periodicity in our data. Further modeling showed that an ARIMA(1,1,1)(1,1,2)[12] improved the base model and resulted in predictions that suggest the need to account for at least a $1 yearly price excursion in natural gas prices owing to seasonality alone.

While it was not initially the purpose of this research, we suggest that future modeling techniques and packages could be developed that further combine machine learning and time series methodology. For example, while it appears that shorter term predictions likely benefit from data-intensive methods such as random forests and machine learning, it would be useful if such methods could be combined more transparently with traditional time series analysis techniques, perhaps by building the NN and RF capabilities in as an additive capability. It may also be desirable to look at ways to combine SARIMA, GARCH, and NN/RF methods given the high degree of heteroskedasticity of the data in areas of interest for forecasting.

In summary, our research into natural gas pricing reinforces the findings of others that continued improvements in modeling financial and price data are possible. Data that may have been assumed to be a random walk may, in fact, contain deeper and previously undiscovered internal and exogenous patterns that would be useful in further predicting and potentially stabilizing markets.

# References

Hancock, E. (2021, February 3). PV-Tech. *Germany’s Federal Network Agency registers 54GW of solar PV.*

<https://www.pv-tech.org/germanys-federal-network-agency-registers-54gw-of-solar-pv/>

Hampton, L. (2021. October 11). Reuters . *U.S. shale producer APA ends flaring, captures more gas as prices soar*. <https://www.reuters.com/business/energy/us-shale-producer-apa-ends-flaring-captures-more-gas-prices-soar-2021-10-11/>

Herrera, G., Constantino, M., Tabak, B., Pistori, H. Su, J-J. Naranpanawa, A. (2019).

*Data on forecasting energy prices using machine learning*. Data in Brief., Volume 25, ISSN 2352-3409. <https://doi.org/10.1016/j.dib.2019.104122>

Long, D., Moore, G., Wenban-Smith, G. (2003). *Fundamentals of the gas market*. Gas Trading Manual. Woodhead Publishing.

Mouchtaris, D. Sofianos E., Gogas, P. & Papadimitriou, T. *Forecasting natural gas spot prices with machine learning*. Energies 2021,14, 5782.

Roberts, D. (2018, July 2).Vox. *Most American homes are still heated with fossil fuels*. It’s time to electrify. <https://www.vox.com/energy-and-environment/2018/6/20/17474124/electrification-natural-gas-furnace-heat-pump>

US Energy Information Administration (2021, October). *Frequently asked questions.* <https://www.eia.gov/tools/faqs/faq.php?id=73&t=11>

US Energy Information Administration. (2021, March) *Electricity in the United States*. <https://www.eia.gov/energyexplained/electricity/electricity-in-the-us.php#:~:text=Natural%20gas%20was%20the%20largest,in%202020%E2%80%94about%2019%25>.

# Appendix A – Reproduceable Markdown Code