Crude Oil as an Explanatory Variable of U.S. Financial Health

Shiqi Cheng, Claire Zhang, Lily Sun, Alina Cai

Massachusetts Institute of Technology, University of Waterloo

February 13, 2024

1 Problem Statement

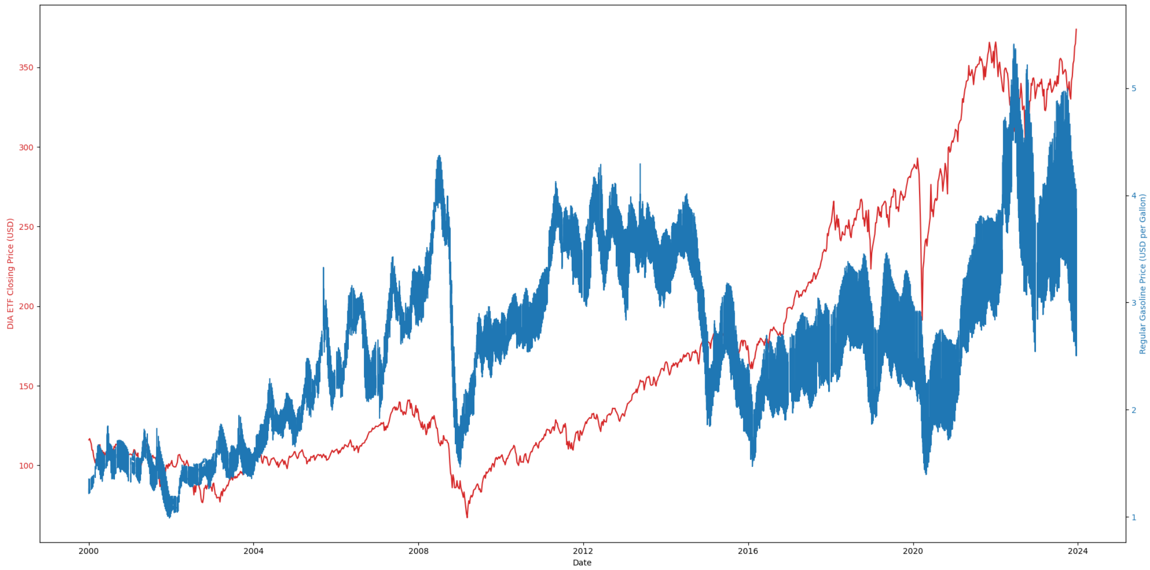


Figure 1: Dow Jones Industrial Average ETF Trust High Prices (Red) and Regular Gasoline Prices (Blue) from 2000-2024.

Changes in gas prices are the value that Americans often believe shows fluctuations in the economy. Oftentimes, consumers assume that high or rising gas prices should lead to concerns about the overall economic health of the United States [1]. The Dow Jones ETF was created to serve as a general stock market and act as an economic indicator. From a preliminary look at Figure 1, Americans would seem to be decently assessing the peaks and pits the market faces through gas prices, especially in large events such as the 2008 financial crisis and COVID-19. However, a thorough analysis of the dataset indicates that crude oil is a much better indicator of the state of the market, and Americans are under the misconception that gas prices is the best indicator of the U.S. economy.

Crude oil is a naturally occurring, unrefined petroleum product composed of hydrocarbon deposits and other organic materials. It serves as a primary energy source globally and is a crucial raw material for a myriad of chemical products, including gasoline (petrol), diesel, jet fuels, and even plastics. The dynamics of crude oil prices significantly influence global economic conditions, given its central role in energy production, transportation, and industrial manufacturing.

The relationship between crude oil and gas prices are inherently strong, primarily because gasoline is derived from the refining of crude oil. As such, the cost of crude oil is a significant component of gasoline pricing, accounting for a substantial portion of the price consumers pay at the pump. The process involves several stages, from crude oil extraction to transportation, refining into gasoline, and distribution to consumers, each adding to the final cost.

2 Executive Summary

What affects the price of stocks? In this report, we analyze the relationship between gasoline price, crude oil price, and their indicative power of the Dow Jones Industrial Average ETF (DIA) (DIA). We have 3 main results

- Crude oil price is more strongly correlated with DIA than gas price is correlated with DIA.
- Our data suggests crude oil price is an explanatory variable in understanding DIA.
- Crude oil supply-side factors have limited impact on crude oil prices.

Research Question: What are the causal mechanisms underlying gasoline price, crude oil price, and the DIA stock price?

2.1 Findings

We began by thoroughly modeling the datasets to understand the dynamic between consumer gas prices, crude oil values, and their implications on the U.S. economy's health, as represented by the Dow Jones Industrial Average ETF (DIA). The genesis of our question stemmed from a common consumer belief that fluctuations in gas prices directly reflect the state of the broader economy.

Our analysis reveals that, contrary to popular belief, crude oil prices are a more accurate measure of economic health than gas prices. This conclusion is supported by data containing crude oil commodity values, gas prices, and the Dow Jones ETF. The relationship between these variables was dissected through correlation tests, revealing crude oil's superior predictive power over the economy compared to gas prices. Additionally, we delved into the intrinsic link between crude oil and gas prices[2]. While gasoline is derived from crude oil, the latter's pricing dynamics play a pivotal role in the global economic landscape beyond the pump. Our models highlights the complexity of these relationships, and shows that the direct impact of crude oil prices on the economy is more pronounced than that of gas prices.

2.2 Implications

This revelation has interesting implications for the policymakers, investors, and the general American public. It Our findings also serve as a call for enhancing economic literacy among consumers, dispelling misconceptions certain indicators being ”tell-all.” In the future, we look to further investigate mechanisms through which crude oil prices influence America’s economic conditions and even global economic conditions and how these insights can be leveraged for more informed decision-making processes.

3 Preliminaries

3.1 Representing Financial Health of the Economy

We represent the financial health of the economy with stock prices. We found that stocks provided in the dataset vary closely with each other and their trends are well represented by the Dow Jones ETF (DIA). This can be seen below in Figure 3 below where the DIA is represented by the bolded gray line.

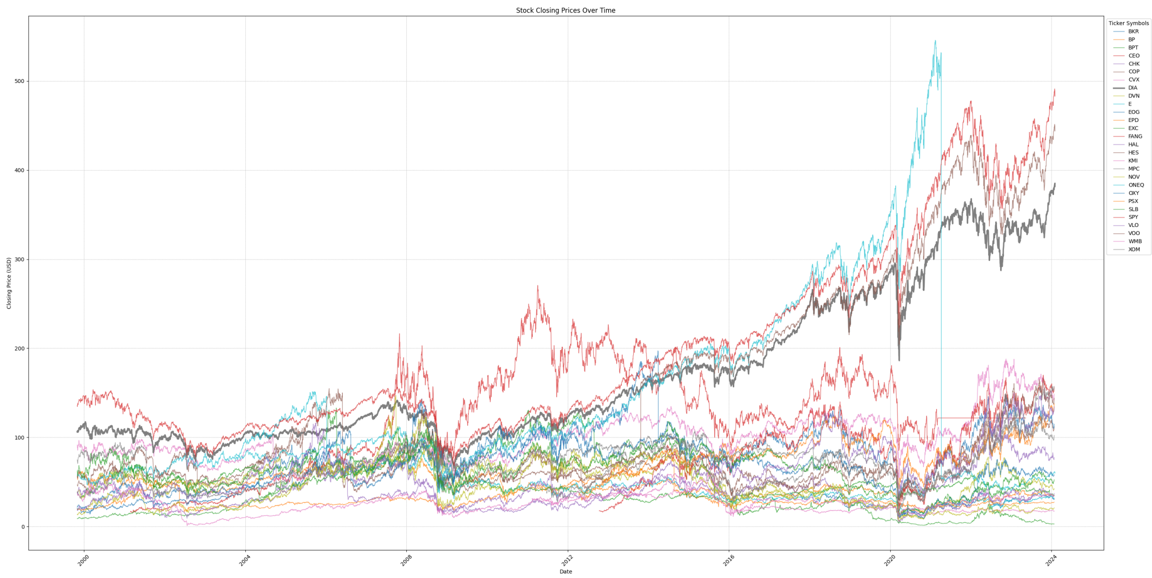


Figure 2: Trends of stock/ETF prices over time.

We ran a correlation test between DIA and the rest of the dataset, and we observed that the DIA is a pivotal indicator of the economy in the U.S. due to its substantial correlation with the S&P 500 ETF Trust (SPY) (0.993507) and Vanguard S&P (0.980674).

Symbol	Correlation
DIA	1.0000
SPY	0.99351
VOO	0.98067
...	...
E	-0.61587
BP	-0.61612
KMI	-0.72110

Figure 3: Correlation Values Between DIA and Other Symbols

S&P 500 serves as one of the main benchmarks of the U.S. equity market and it helps denote the financial health and stability of the economy, so we can safely assume that DIA could be used as a representative to measure the economy [3]. While we can also noticed some lower correlations, these were often with smaller, individual companies that experience specific risks and industrial influences. Overall, we can see that the DIA offers an accurate perspective of the U.S. market and some of its largest contributors.

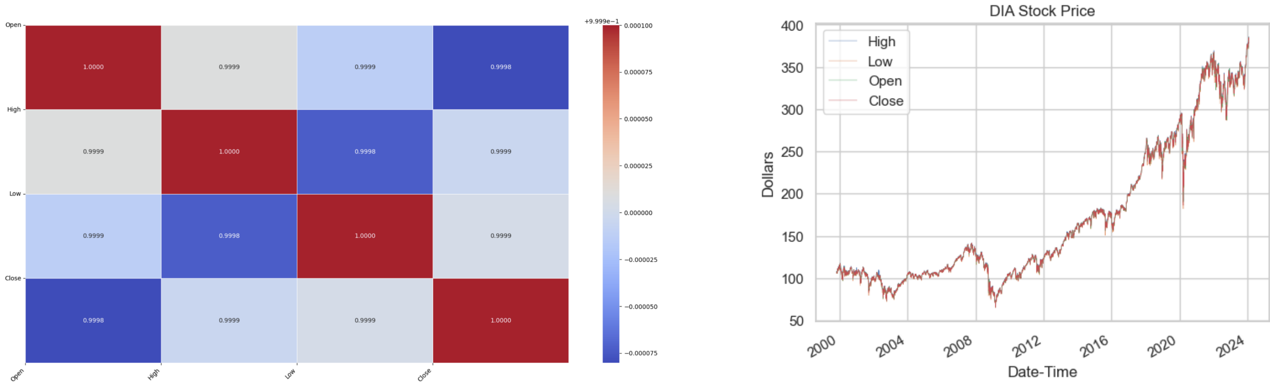
For the rest of our analysis and plot labels, we use DIA to refer to the general stock price as an indicator of financial health.

3.2 Invariance of Stock Types

Based on the stock and ETF data which offers four different parameter ("Open", "High", "Low", "Close"), we noticed that the correlation between the four parameters is high (between 0.9998 and 1.000) as seen in Figure 4. For simplicity, we chose the"High" parameter to represent stock values in the rest of our data analysis.

3.3 Data

We clean and filter data from `all_commodities.csv`, `all_stocks_and_etfs.csv`, and `weekly_gasoline_prices.csv`. We aggregate data by week, spanning 2000-01-09 to 2023-12-31, as this is the range where our targeted data has all non-zero values.



(a) Heatmap representing the correlations between the four different stock measures

(b) Trends of the four stock measures over time

Figure 4: The four stock prices move closely together.

4 Model

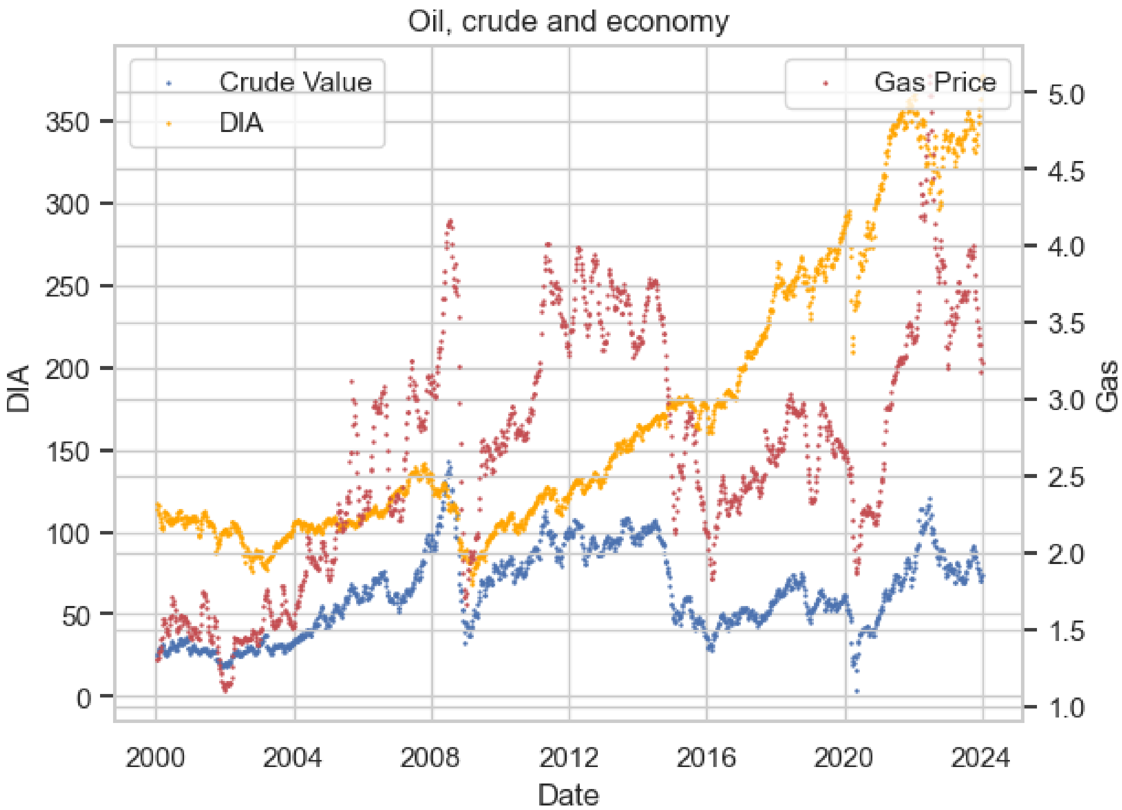


Figure 5: Trends of Gas, Crude Oil, and the Economy over time

4.1 Explaining DIA

We sought to model the factors and mechanisms affecting DIA. We initially analyze the relationship between gas price and DIA.

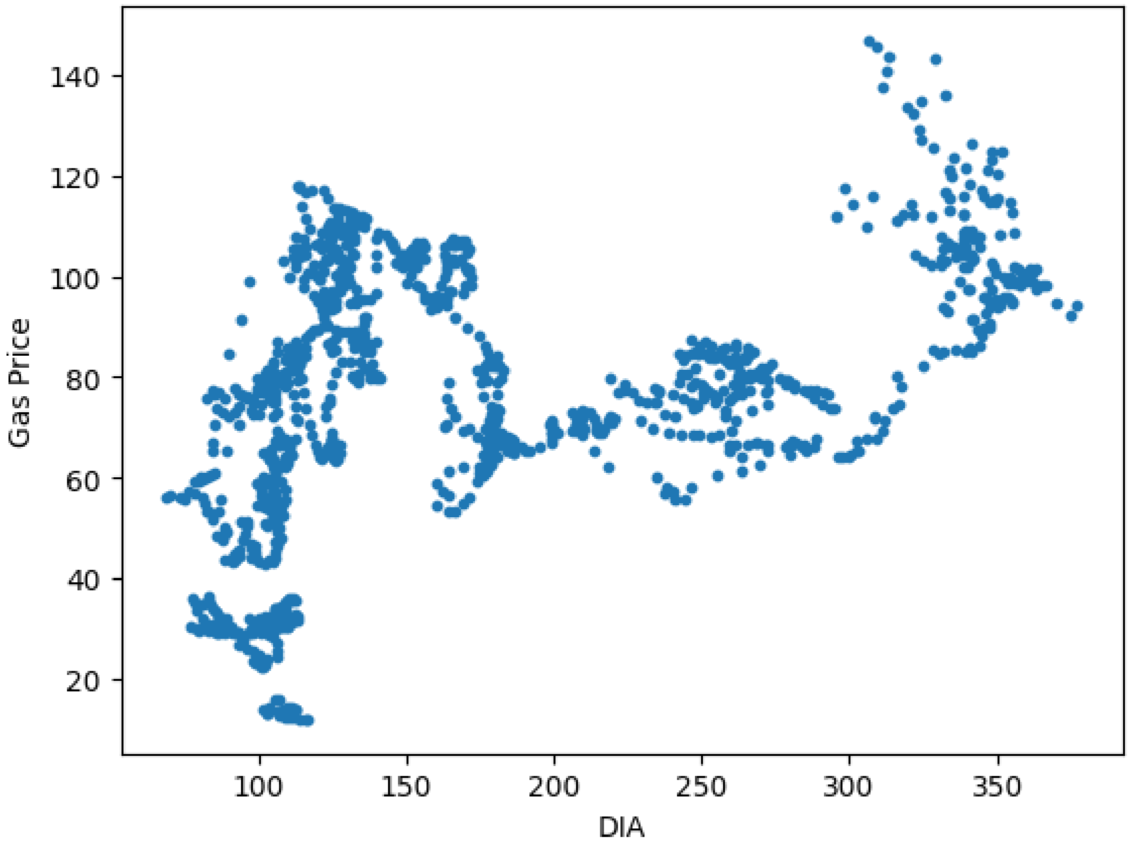
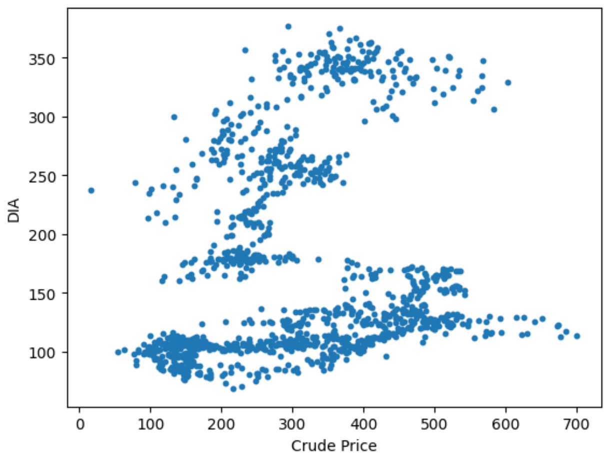
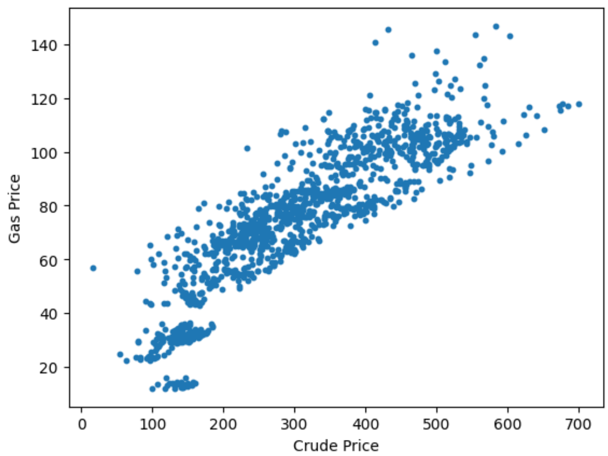


Figure 6: DIA vs. Gas Price over Time

Although DIA vs. gas attain a Pearson correlation coefficient of 0.48, their relationship is not linear. As crude price is a determining factor of gas price, we then analyze the effect of crude price on DIA.



(a) DIA vs. Crude Price over Time

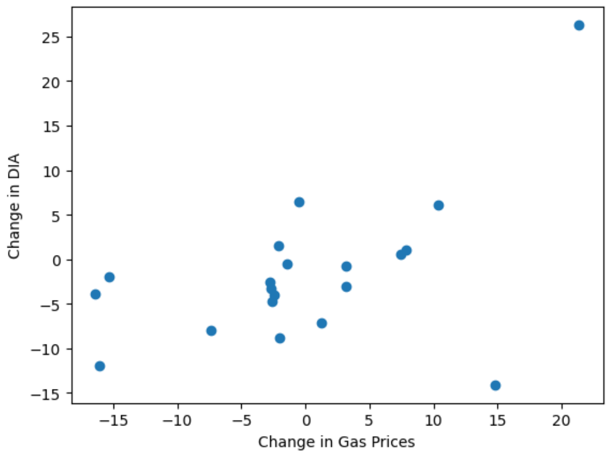


(b) Gas vs. Crude Price over Time

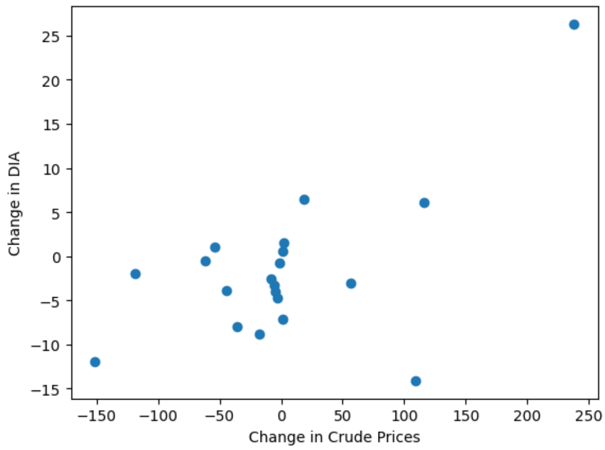
Figure 7

We observe a mixture of linear relationships, which indicates that there is a more direct relationship between crude price and DIA. Figure 5 indicates correlation of *changes* in DIA with *changes* in crude and gas prices.

We thus investigate the relationship between short-range fluctuations in DIA and prices. We calculate short-range fluctuations of a week as the average of next 8 weeks' price minus the average of the last 8 weeks' price.



(a) Change in DIA and Gas Prices over Time



(b) Change in Gas and Crude Prices over Time

Figure 8

DIA vs. change in crude price has a correlation of 0.61 while DIA vs. change in gas price admit a lower correlation of 0.517. The relationship is linear as seen above. We model DIA, change in gas price, and change in crude price as a causal model, where we assume i.i.d. observations, linear mechanisms, and additive noise.

$$X_{\text{DIA}} = w_{\text{DIA}}X + \epsilon_{\text{DIA, Crude}}$$

$$X_{\text{Crude}} = w_{\text{Crude}}X + \epsilon_{\text{Crude, DIA}}$$

$$X_{\text{Gas}} = w_{\text{Gas}}X + \epsilon_{\text{Gas, Crude}}$$

These variables can be modelled as a complete graph, whose underlying causal model is a DAG. We learn the weights w through NOTEARS, an analytical algorithm that solves $X = WX$ under a acyclicty constraint. [4]

We contrast this with the learned weights (NOTEARS algorithms, same methodology) of absolute prices and DIA. We conclude that crude prices are an explanatory variable between the apparent link between gas prices and DIA.

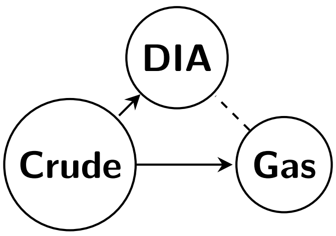
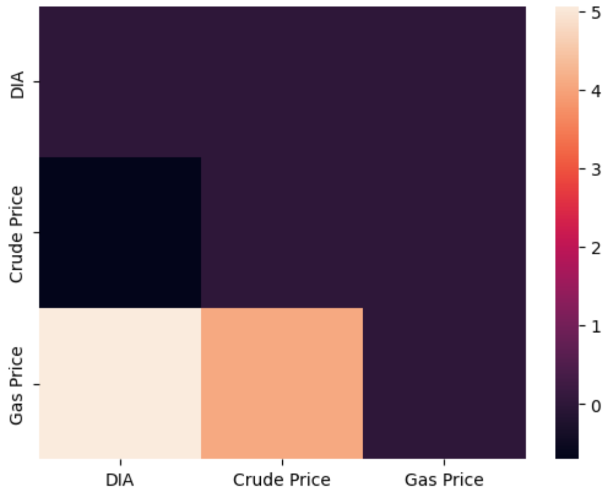
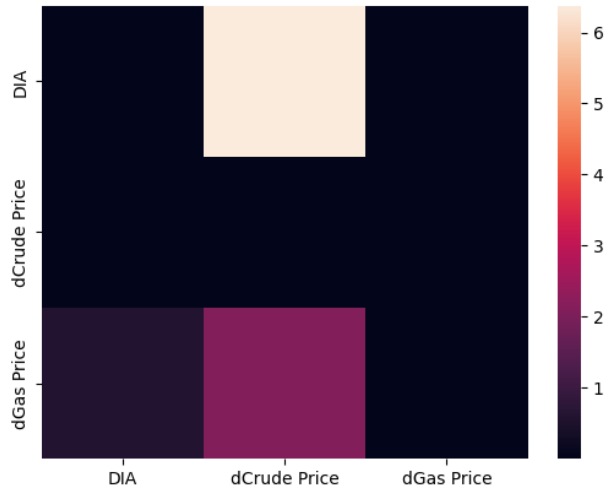


Figure 10: Fluctuation mechanisms



(a) Weights between DIA and price fluctuations



(b) Weights between DIA and absolute price

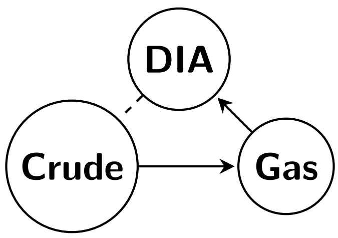


Figure 11: Absolute mechanisms

4.2 Small effect of crude oil supply side factors

We saw the strong correlation between crude oil prices and DIA. Thus, we hypothesised that crude oil prices is affected more by the outlook and state of the economy, rather than supply-side factors such as the amount of crude oil harvested.

Analysing the supply-side data, we look at domestic field production, imports and exports of crude oil, and how they relate crude oil prices.

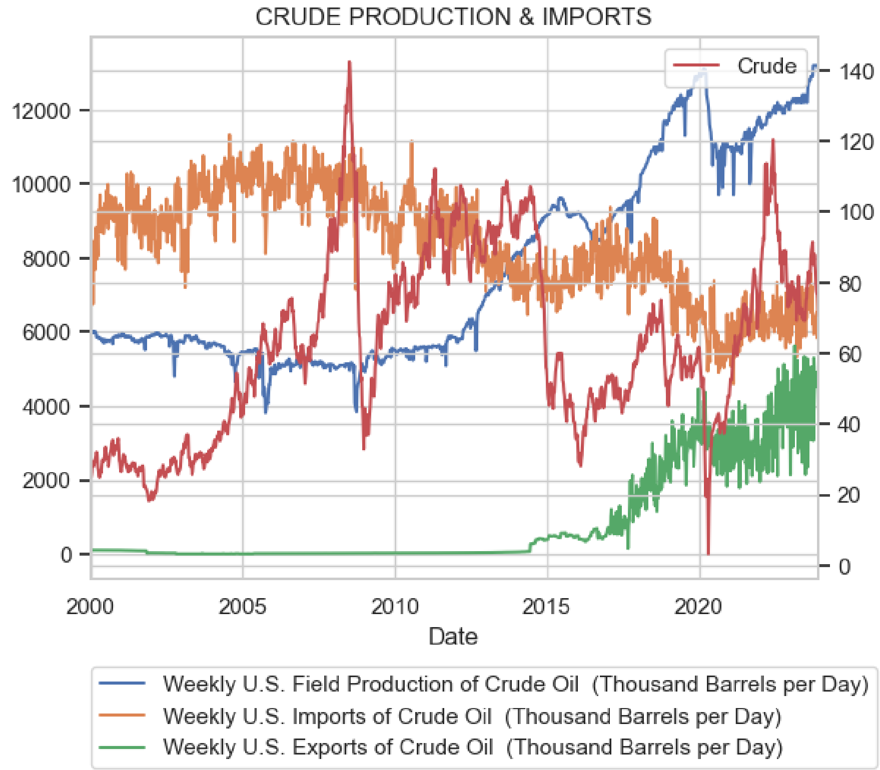


Figure 12: Crude oil prices (red) and amounts produced/imported/exported over time

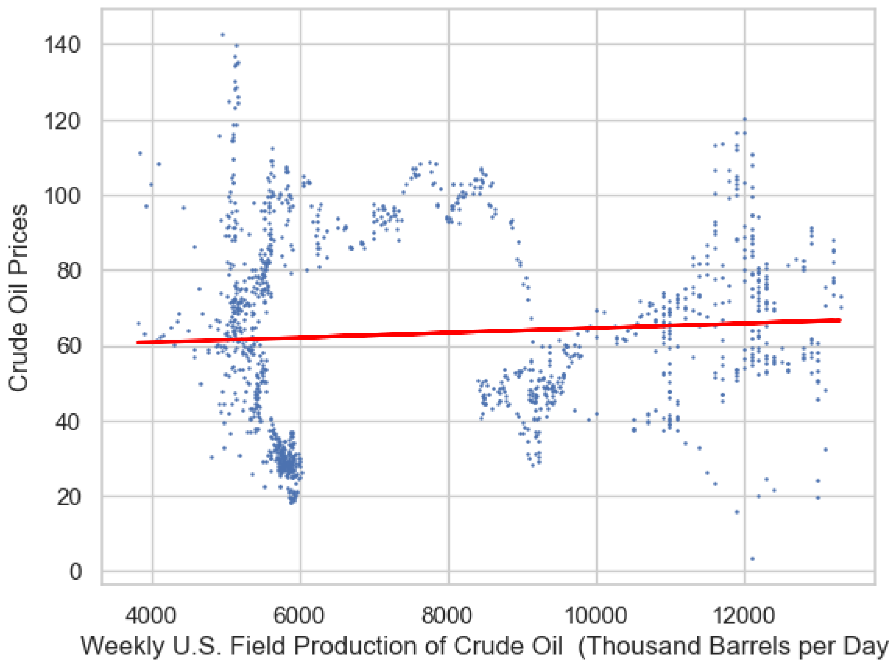


Figure 13: Graph of crude oil prices against domestic production. r = 0.0655

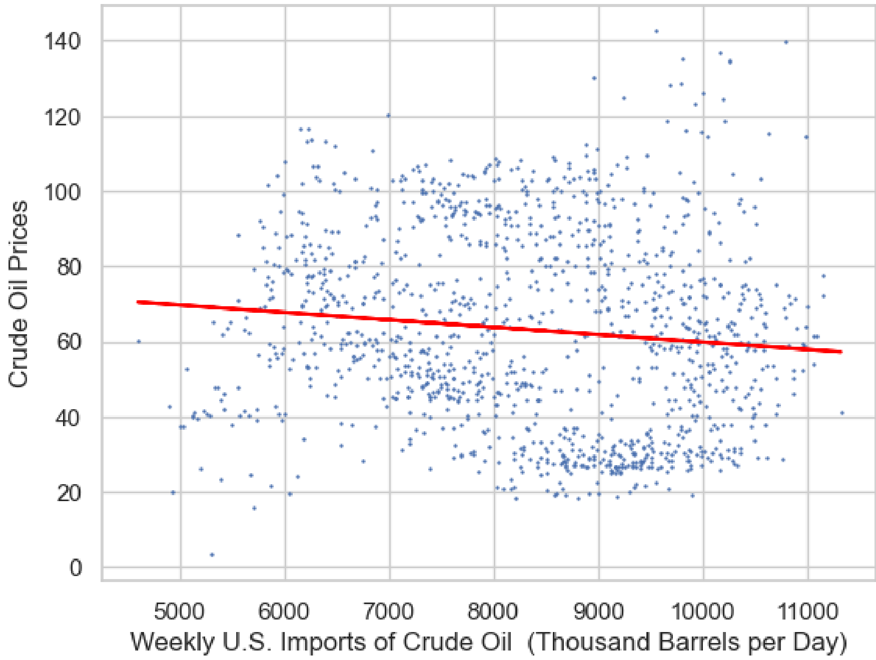


Figure 14: Graph of crude oil prices against imports. r = -0.109

If supply-side factors were important, we would expect both domestic production of crude oil and imports of crude oil to correlate with prices, given the US imports about 40 percent of crude oil processed. However we see that is not true. This could be because crude oil prices are largely determined in the futures market, which generally depends on the state of the economy and economic outlook, rather than its supply. Furthermore, since crude oil is traded internationally, there is a lot of noise in its price from external factors such as other countries' demand, economies and exports, which we do not have the data for, and so domestic production is only a small piece of the global crude oil market.

4.3 Small effect of demand-side factors on gas prices

Our initial hypothesis was that the largest influence on gas prices was consumer demand. If this were true, we would have seen a strong correlation between the amounts consumers drive and gas prices.

After a careful analysis of the transportation dataset, we instead see a clearer seasonal pattern in the amount consumers drive. We believe this is due to the general seasonal variation in amounts people travel, as seen in both the 'miles driven' and 'railway use' graphs, that is greater in warmer months and holiday seasons. This seasonal pattern is absent in gas prices, showing that gas prices and transportation are weakly correlated.

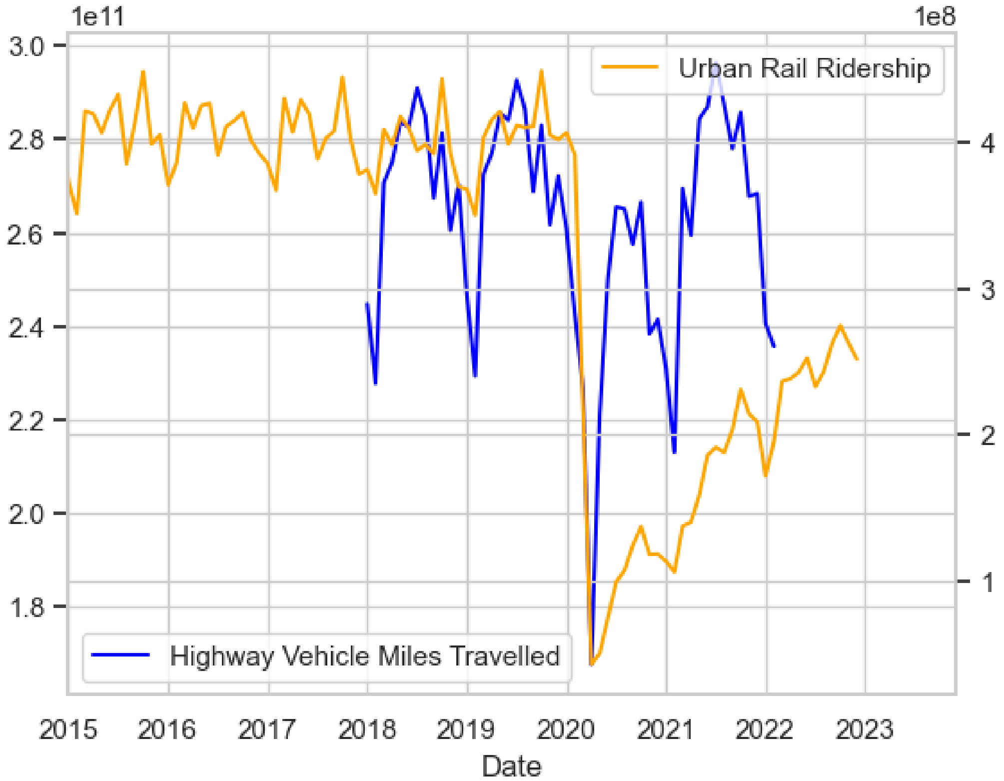


Figure 15: Seasonal variation in distances driven and rail usage over time

5 Appendix

5.1 Future Research Areas

The findings of our study suggest several promising directions for future research. One area involves examining the causal mechanisms between crude oil prices and specific sectors of the economy to better understand their differential impacts. Additionally, investigating the role of renewable energy sources and their pricing on the economy could provide insights into the transition towards a more sustainable energy landscape, assisting the prominent global issue of climate change. Further, the exploration of predictive models incorporating machine learning techniques to forecast economic health based on energy prices could enhance the precision of economic indicators. Lastly, a comparative analysis of global markets could potentially lead to a more comprehensive understanding between energy prices and economic health on a global scale.

5.2 Improvements on Methodology

The NOTEARS learned model for fluctuations has a acyclicity loss of $5.77 \cdot 10^{-9}$ and least squares loss 872.64. The high least squares loss suggests that not all pairs among fluctuation in gas price, fluctuation in crude price, and DIA are linear. DIA vs. gas price appears linear (Figure 7(a)). Problematically, the learned DAG shows DIA causes (fluctuations in) prices, not the other way around.

This is inconsistent with intuition and further analysis should be done to either suggest that the arrow is not strongly directed, or that we have ignored other causal variables.

5.3 Unsuccessful Analysis Pathways

We originally looked at volatility of crude prices, oil prices and DIA. Volatility here is calculated by the standard deviation over a moving window of $\log(\%\text{ change in price})$. An example of volatility over a window of 10 weeks is given as follows:

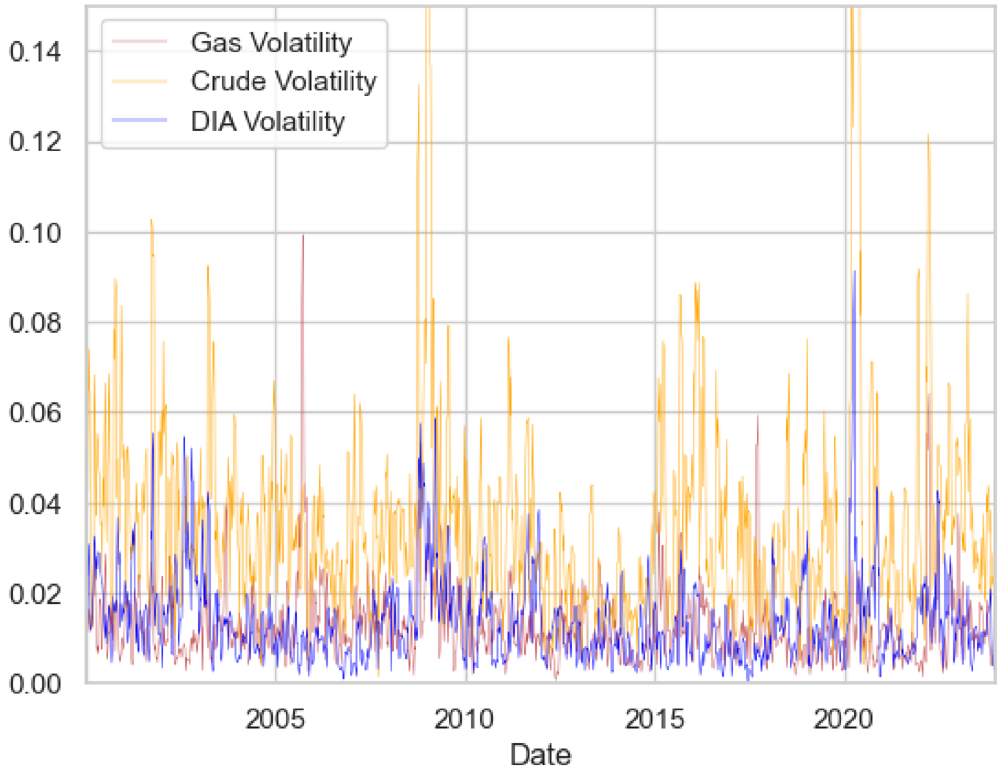


Figure 16: Volatility of prices over time

The correlation coefficients between the volatilities are given as follows: crude and gas 0.208, crude and economy 0.402, gas and economy 0.173.

References

[1] Edelstein, P., and Kilian, L. Do lower gasoline prices make for confident consumers? *Pew Research Journal* (2007).

[2] EIA. Oil and petroleum products explained. *ArXiv* (2023).

[3] Quandt, A. K. Stock market correlations to economic indicators. *University of Nebraska-Lincoln* (2020).

[4] Xun Zheng, Bryon Aragam, P. R., and Xing, E. P. Dags with no tears: Continuous optimization for structure learning. *ArXiv* (2018).