1. Forecast thesis

This paper forecasts U.S. gasoline prices for the end of November 2025 using a multivariate time-series regression model. The model incorporates core market fundamentals, such as copper futures prices, Brent crack spread, the crude roll spread, and U.S.-euro exchange rate. By examining the dynamic relationship between these factors and monthly gasoline prices, the forecast aims to assess whether current production trends and refining margins point to upward or downward pressure on consumer fuel prices in the coming month. Understanding the direction of gasoline prices is critical for policymakers and investors, as it reflects broader energy market conditions and directly impacts inflation and consumer spending.

2. Data sources and methodology

2.1 Data

The data was primarily collected from the U.S. Energy Information Administration (EIA), covering key variables related to U.S. energy production, inventories, and consumption from January 2020 through October 2025. Additional macroeconomic indicators such as the Consumer Price Index (CPI), Gross Domestic Product (GDP), and unemployment rates were retrieved from the Federal Reserve Economic Data (FRED). As well as futures data for commodities closely tied to the energy market such as crude oil, soybean oil, copper and gold were obtained through LSEG (London Stock Exchange Group).

Each feature was aligned to a monthly frequency to ensure temporal consistency across different data sources, which originally included weekly, monthly, and annual observations. Missing values were addressed using a linear interpolation, ensuring a smooth transition between data points without introducing discontinuities. The period from March to July 2020 was excluded to remove distortions caused by extreme volatility and structural breaks associated with the Covid-19 pandemic. Data is available here.

2.2 Methodology

2.2.i Variable Selection

Out of roughly 50 predictors, we used the Lévy area to find variables that consistently moved before US gas prices. The Lévy area between two once-differenced time series tells us about their timing relationship. When X = US Gas Prices, a negative Lévy area means the Y variable leads X. We calculated the Lévy areas using a version of the shoelace polygon theorem adapted for discrete data points:

$$A^n = \frac{1}{2}\sum_{i=0}^{n-1}(x_i y_{i+1} + y_i x_{i+1})$$

Since the resulting Lévy areas have units of $X_{units} \cdot Y_{units}$, which isn't easy to interpret, we made the areas dimensionless by dividing by the product of X and Y's standard deviations:

$$A^{n'} = \frac{A^n}{\sigma_X \cdot \sigma_Y}$$

To ensure our results weren't due to chance, we used a block resampling method to calculate p-values. This method shuffles blocks of data to break up long-term patterns while keeping short-term relationships intact. We recalculated the Lévy area for each shuffled version and used it to compute the p-value of the real value.

We expected that gas price drivers vary by season (peak demand in summer, lowest in winter, transition periods in spring and fall), so we only used data from the demand season we are predicting (winter). We selected variables that met three criteria: negative Lévy area, p_value <= 0.10, and data with updated values available for the date of October 30th. The variables that met these conditions were 1) front-month copper futures price, 2) Brent oil crack spread 3) WTI intermediate front-month / spot spread and 4) USD / Euro exchange rate.

	Var	Levy Area	p-value
0	copper_fut_price	-11.10	0.075
1	brent_crack_spread	-15.05	0.027
2	crude_roll_spread	-7.09	0.005
3	us_euro_exchg	-13.14	0.054

3. Model outputs and interpretation

3.1 Model fitting and coefficient interpretation

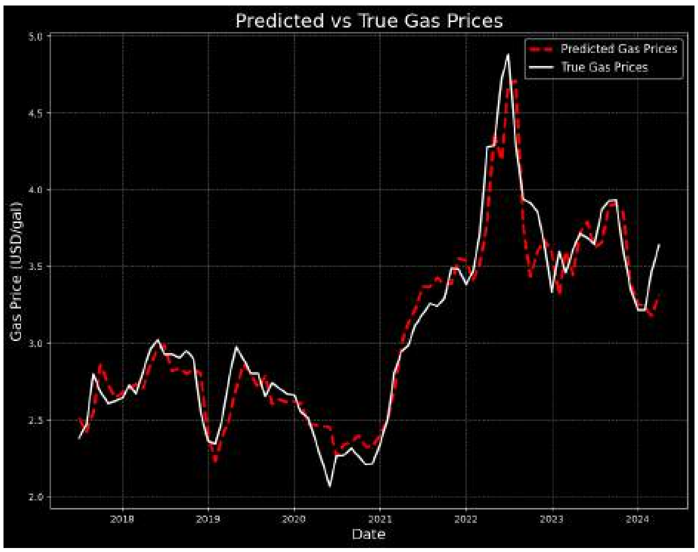
We fit a multivariable regression on the selected variables, and also included the current price of gas as an autoregressive variable. The resulting model achieved an R^2 score of .907

Variable	β_i	t	$p > \|t\|$
Copper Future Price	0.1490	3.312	0.001
Brent Crack Spread	0.0055	2.061	0.042
Crude Roll Spread	0.1525	1.936	0.055
US/Euro Exchange Rate	0.2152	1.219	0.225
Current gas price	0.798	14.858	0.000

*the constant term was not included for space purposes

Current gas prices were by far the strongest indicator, with about 80% of the previous month's value carrying over into the next and indicating strong price persistence. As a result, the regression essentially provides a directional forecast of current prices, adjusted by market fundamentals. The positive copper coefficient makes sense since copper acts as an industrial barometer and rises with broader demand. The positive Brent crack spread indicates that when refinery margins expand due to higher wholesale gasoline prices, retailers pass on those costs to consumers. A positive crude roll spread reflects backwardation, tight current supply and strong demand, so wider spreads correspond to higher prices. The U.S.-Euro exchange rate was the weakest predictor, but was retained to avoid post-hoc variable selection.

3.2 Results



Avg. out of sample test error: 13.3 cents

3.3 Prediction of November Month-End Value
Current values:

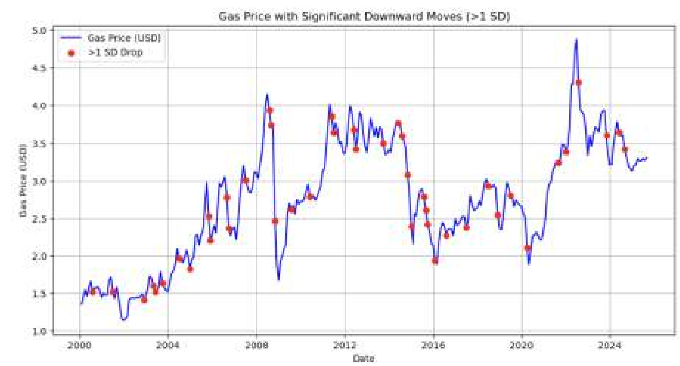
Copper Futures ($/lb)	Brent Crack Spread ($/bl)	Crude Roll Spread ($)	US/EU Exchg	Gas ($/gal)
5.10	18.0	-0.16	1.16	3.038

Predicted November month-end gas price: $3.35 / gal

4. Risks and alternative scenarios

Our model will likely fail to capture the impact of major headlines or unexpected shocks that drive short-term volatility. For instance, the recent announcement by OPEC indicating a potential output increase in December led to a global drop in crude oil prices, which in turn lowered U.S. gasoline prices. In addition, the data used for training ends about one month before the forecast period, which means the model may miss recent market dynamics or geopolitical developments, such as the one mentioned above, that can significantly influence gasoline prices.

To address these limitations, we examined how often gasoline prices fell by more than one standard deviation relative to their percent change. This analysis only focuses on downward movements because our outlook is based on maintaining a price over $3.00 per gallon; upward deviations would be favorable for our forecast. Using a six-month rolling window to capture short-term trends, we found that prices dipped more than one standard deviation below the mean 13.9% of the time. This means an 86.1% likelihood that prices stay above our target. To further illustrate, we computed the price associated with a two-standard-deviation drop from the most recent observed price of $3.038. Using the same methodology, this price would be $3.014, reinforcing our confidence in predicting that gasoline prices are likely to remain above $3.00 per gallon.



Months with over 1 SD drop: 13.92%

5. Final probability

Combining our model's predicted price of $3.35 per gallon, which exceeds the $3.00 threshold, with the historical analysis showing an 86.1% likelihood that prices do not fall more than one standard deviation below the mean, and noting that a one-standard-deviation drop from the recent price of $3.038 would still be $3.014 (above $3.00), we confidently predict that U.S. gas prices will finish above $3.00 per gallon.