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