Squeezing the Pump

A Threshold-Regime Analysis of U.S. Refinery Utilization
and Its Impact on Retail Gasoline Price Dynamics

Team #10

Zhixuan Liu, Zeynep Nazli Takci, Katrin Schwab, Ngoc Linh Pham

May 2025

Contents

1. Introduction

Retail gasoline prices in the United States are driven by a complex interplay of supply-side factors, among which refinery utilization plays a pivotal role. Yet, the sensitivity of pump prices to changes in refinery capacity is not uniform: periods of high utilization ("stress" regimes) exhibit markedly different price dynamics than more "easy" operating conditions. In this report, we investigate this regime-dependent behavior using weekly U.S. data from the Energy Information Administration—refinery utilization, regular-grade pump prices, WTI spot prices, and gasoline ending stocks—over a 15-year window (2010–2025).

We first apply Hansen's sup-LM threshold test to bi-weekly price changes and refinery utilization, pinpointing a structural break at 84.5%. Rolling-correlation analyses and separate OLS regressions quantify regime-wise slopes—revealing a slight positive sensitivity below the threshold versus a stronger negative response above it. Building on this, we show that a simple logistic spike-prediction model augmented with a binary "Stress" flag boosts ROC-AUC from 0.69 to 0.74 and doubles Precision@100. Finally, we propose a "Buying the Dip" regime cross-over trading strategy -- entering long when utilization dips below 84.5% and reversing to short on its rebound -- and demonstrate its resilience to drawdowns in an out-of-sample backtest.

The report is structured as follows:

- **Section 2** details data sources and preprocessing.
- **Section 3** presents the exploratory regime-break analysis.
- **Section 4** develops and evaluates the spike-prediction model.
- **Section 5** introduces the regime cross-over trading strategy.
- **Section 6** concludes with key insights, practical implications, and avenues for future work.

2. Data & Preprocessing

2.1 Data Sources and series

- Refinery utilisation: Weekly U.S. percent of operable capacity (series WPULEUS3), sourced from the U.S. Energy Information Administration.
- Retail gasoline price: Weekly U.S. regular-grade pump price in $/gal (series EMM_EPMR_PTE_NUS_DPG), U.S. EIA.
- WTI spot price: Weekly West Texas Intermediate closing price in $/bbl (series RWTC), U.S. EIA.
- Gasoline ending stocks: Weekly U.S. total gasoline stocks in thousand barrels, U.S. EIA.

2.2 Data sampling and processing

The sampling window is taken over the past 15 years from 01-01-2010 to 01-01-2025, with the most recent volatile months deliberately excluded to ensure a stable regime-break analysis. All the time series data files are downloaded directly from the EIA website which contains a “Date” column and the relevant variables of interest. Misalignment of data recorded date are handled by converting each given date into a “Week” period ending on Sunday. Missing values are removed.

3. Regime-Break Analysis

3.1 Hansen’s sup-LM threshold test

We measure the impact of today’s refinery utilisation on retail prices two weeks ahead by defining

$$\Delta P_t^{(2)} = P_{t+2} - Pt$$

in cents per gallon. Statistical analysis demonstrates about half of the change in crude oil price is passed through to retail prices within two weeks of the price change, all other market factors equal. Demand for crude oil is very highly related to the demand for refined products [1].

Figure 1 shows the scatter plot of $\Delta P_t^{(2)}$ vs Utilisation U_t, hinting at two distinct behaviors. When utilisation is below roughly 84–85 %, price changes are muted and roughly flat; while as the percent of utilisation reaches above 90%, price drops have become more noticeable as utilisation increases. Clearly, there are two distinct regimes: when utilization is below roughly 84–85% and when it exceeds that threshold. Thus, a standard linear fit would average over both, obscuring the subtle regime shift.

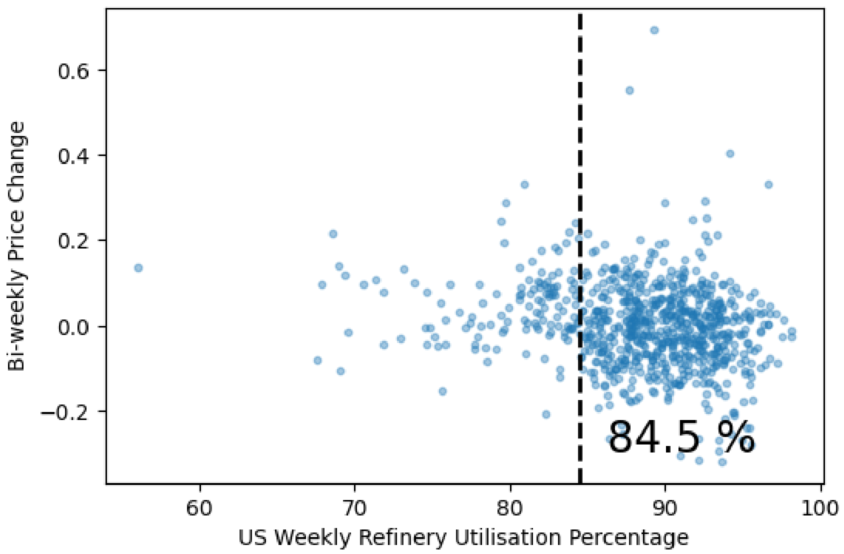

Figure 1 Scatter plot of bi-weekly Price Change vs Utilisation

A Hansen's sup-LM threshold test is thus designed to objectively identify the regime threshold at which refinery utilisation shifts price-sensitivity. In this statistical testing framework, two models are compared:

1) The single-slope model

$$\Delta P_t^{(2)} = \alpha + \beta U_t + \epsilon_t, \quad \epsilon_t \sim i.i.d$$

with residual sum of squares RSS_0.

2) The two-regime model

$$\Delta P_t^{(2)} = \alpha_1 + \beta_1 U_t \cdot 1\{U_t \leq c\} + \alpha_2 + \beta_2 x_t \cdot 1\{U_t > c\} + \eta_t$$

where $\mathbf{1}\{\cdot\}$ is the indicator function and RSS_1 is the corresponding residual sum of squares.

The sup-LM statistic measures the maximum Lagrange-multiplier test value across all candidate threshold values, quantifying the strongest evidence against a single-slope (no-break) model in favor of a two-regime specification. Let T be the number of observations and let p and $p + q$ be the number of parameters in the single-slope and two-regime models, respectively (so q is the number of extra parameters, in our care $q = 1$). The sup-LM statistic is then defined as:

$$\sup LM = max\, LM(c) = \frac{(RSS_0 - RSS_1(c))/q}{RSS_1(c)/(T - P)} \approx \frac{T(RSS_0 - RSS_1(c))}{RSS_1(c)}$$

$$c^* = \arg max\, LM(c)$$

The threshold the divides two regimes c^* is therefore obtained at $\sup LM$. A wild boostrap (Rademacher draws on the restricted-model residuals) is also performed here to assess significance. The statistical test results are show in table 1.

Estimated threshold c^*	sup-LM statistic	bootstrap p-value	95% CI for c*
84.5%	8.63	0.064	[84.2%, 85.4%]

Table 1 sup-LM test statistics

It is important to state here that several parameters, including look-back period and price-shift windows, have been tested to understand the robustness of the Hansen's sup-LM threshold test results. The value of $c^* = 84.5\%$ has been achieved consistently thought these hyperparameter sets.

3.2 The 52-week rolling correlation

The 52-week rolling correlation highlights how the relationship between utilisation and subsequent price changes evolves over an entire year. Short-term noise can be smoothed out while shifts are still identified in the underlying dynamics. Table 1 presents the year-long rolling correlations between refinery utilisation and two-week price changes, split by regime.

	mean	std	min	max
Easy regime (Util ≤ 84.5%)	-0.240769	0.221478	-0.535967	0.366539
Stress regime (Util > 84.5%)	-0.215279	0.199851	-0.652475	0.383404

Table 2 52-week rolling statistics for two regimes

In the "easy" state (≤ 84.5 %), utilisation and prices tend to move modestly in opposite directions, whereas once utilisation exceeds that break the relationship weakens and becomes far more unpredictable. This confirms a clear structural shift at the threshold and validates using a simple Stress flag to anticipate volatile price moves.

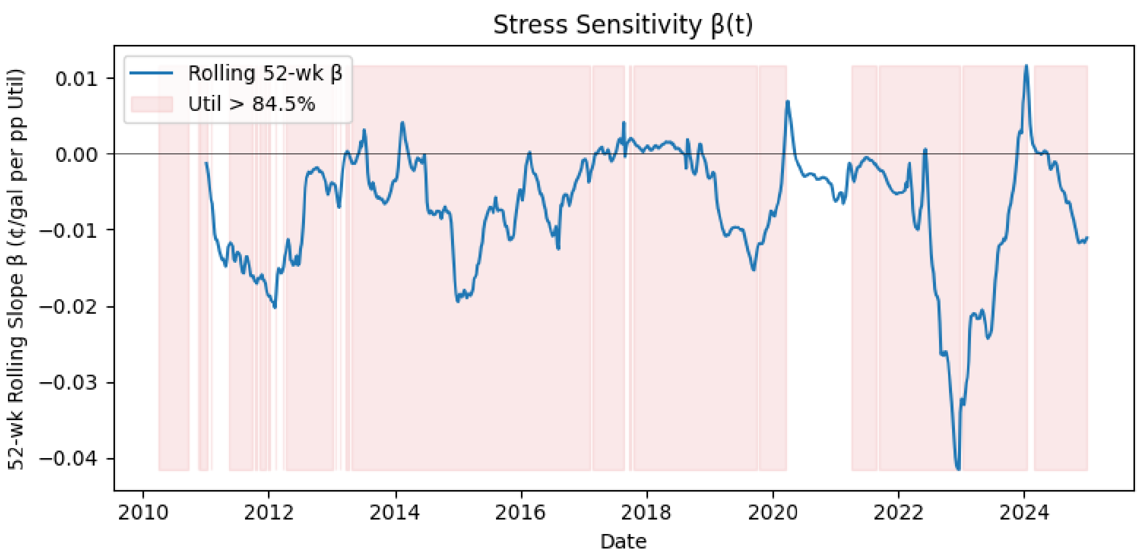

Figure 2 52-week rolling slope of bi-weekly price change vs utilisation

The following figure is plotted by fitting a simple OLS (single-slope model) of the two-week forward price change on utilisation within a rolling 52-week. Shaded regions represent the Stress regime. The long-run sensitivity of bi-weekly price β(t) evolves over time and differs between “easy” and “stress” regimes. Consistently negative slopes in shaded Stress regime show that higher utilisation pulls prices down over the next two weeks, indicating rapid nationwide supply relief.

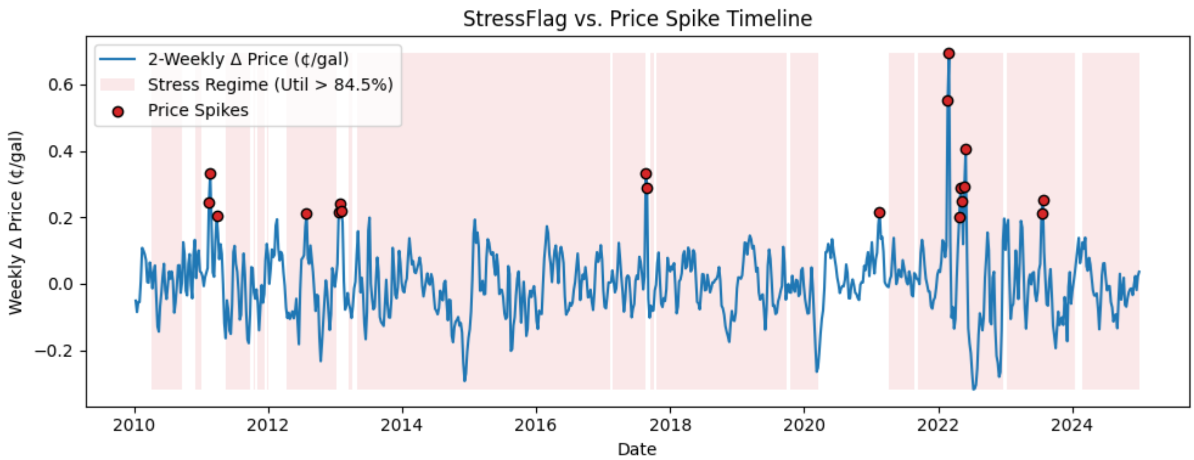

Figure 3 Price Spike markings

Figure X shows the bi-weekly change in retail gasoline prices (blue line) overlaid with shaded periods where refinery utilisation exceeds 84.5 % and red markers for weeks in which $\Delta P_t^{(2)}$ exceeds 2 standard deviation from its mean. The clustering of nearly all marked price spikes within these shaded Stress periods confirms that high-utilisation regimes reliably precede large upward price moves.

3.3 Regime-wise sensitivity:

Analyses in the previous section has prompt us to investigate the regime-wise sensitivities of retail gasoline price to percentage utilisation of nationwide refineries. Regular OLS regression is used to estimate the two slopes β_1 and β_2 appeared in the two-regime model equation. The fitted results are shown below.

	mean $\Delta P_t^{(2)}$ ¢	std $\Delta P_t^{(2)}$ ¢	Fitted slope β (¢/pp)	β_2/β_1	Welch t-test	p-value
Easy regime (Util ≤ 84.5%)	+0.048	0.086	+0.00092	-3	6.74	1.5×10^{-10}
Stress regime (Util > 84.5%)	–0.009	0.101	–0.00279			

Table 3 OLS regression results for two regimes

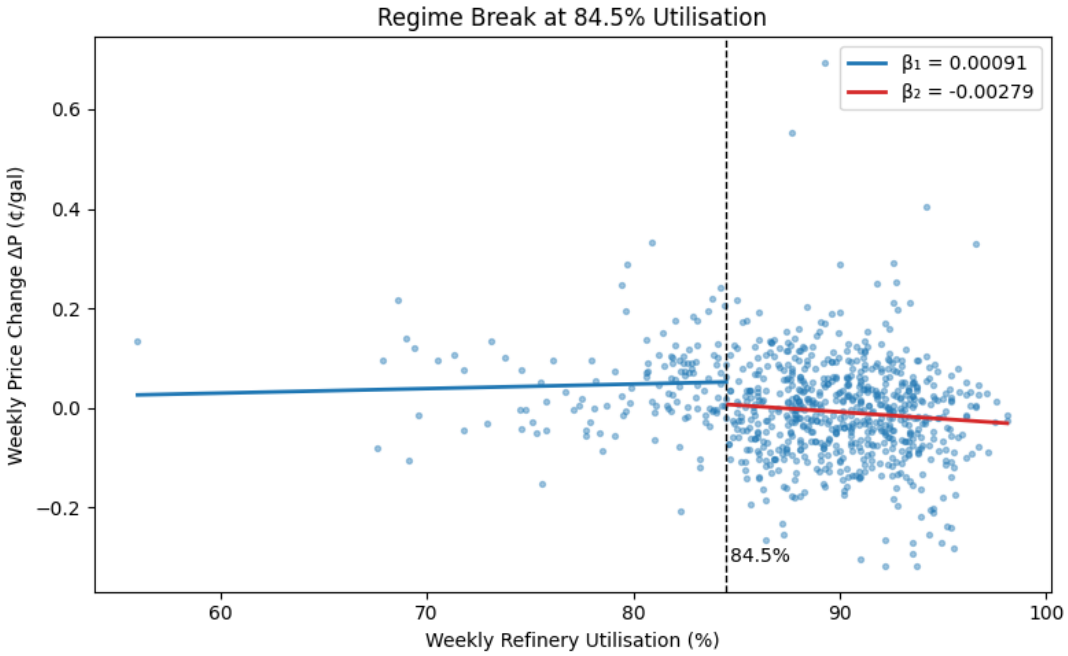

Figure 4 OLS regression results for two regimes

The OLS estimates reveal a clear sign reversal across regimes: in the low-stress state, the average two-week price change is slightly positive, whereas in the high-stress the mean price change is slightly negative, 3× larger magnitude but opposite direction compared to the easy regime. A Welch two-sample t-test confirms that the difference in mean $\Delta P_t^{(2)}$ between the two regimes is highly significant, confirming the validity of the 84.5 % break as a structurally distinct price-sensitivity threshold.

4. Spike prediction model

We evaluate whether adding the Stress flag to a simple spike-prediction model measurably improves early-warning performance. Our target variable is a binary spike indicator defined as

$$\text{spike}_{\,t} = \mathbf{1}\left(\Delta P_t^{(2)} > \mu_{\Delta P^{(2)}} + 2\sigma_{\Delta P^{(2)}}\right)$$

This indicator marks weeks in which the two-week price change exceeds its historical mean by two standard deviations.

As a baseline, we fit a logistic-regression model using only week-over-week WTI changes $dWTI_t$ and gasoline ending stocks changes $dStock_t$. Both models are estimated via maximum-likelihood (`scikit-learn LogisticRegression` with `liblinear` solver). We compare their out-of-sample ranking performance using ROC-AUC and Precision@100 on the full sample.

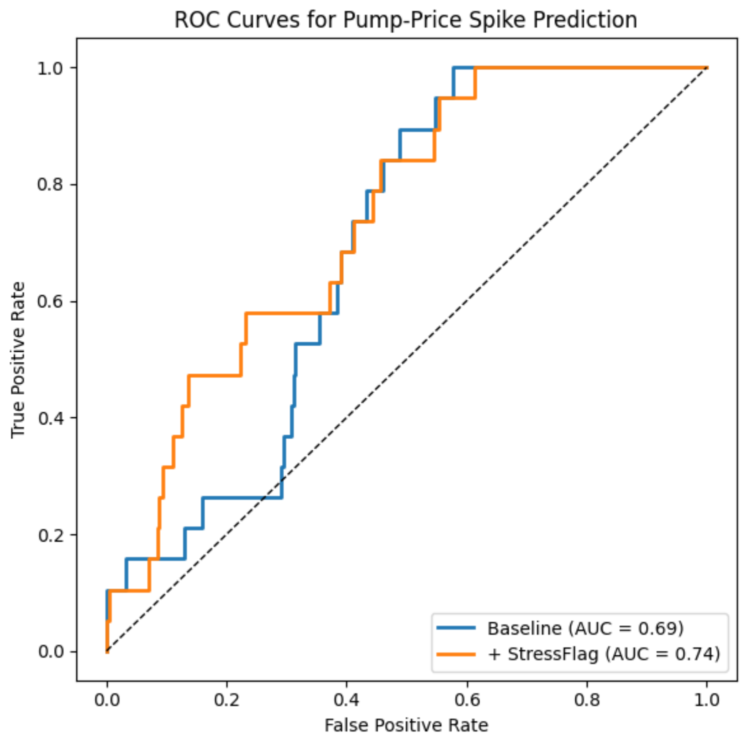

Figure 5 Spike prediction ROC curve

Model	ROC-AUC	Precision@100
Baseline (WTI+stocks)	0.69	3%
Baseline + StressFlag	0.74	7%

Table 4 Spike prediction metrics for two models

The augmented model that includes our Stress flag clearly outperforms the baseline. Adding the binary utilisation indicator raises the ROC-AUC from 0.69 to 0.74, and more than doubles Precision@100.

These gains confirm that refinery stress provides independent predictive power beyond crude-price and inventory swings alone. In practice, improving precision by 4 pp means cutting false alarms by over half for a fixed alert budget.

5. Regime Cross-Over Trading Strategy

5.1 Strategy description

The previous findings show the opposite sensitivity of retail gasoline price to in different regimes Percent Utilization of Refinery Operable Capacity. Taking advantage of this difference could potentially result in a statistical-signal-driven trading strategy. In this section, we will introduce such a strategy – *"Buying the Dip"*.

A dip is a structure in the utilisation curve highlighted in the figure below, where utilisation percentage U_t first drops below the regime threshold c^*, entering the Easy regime; and within 4 weeks, utilisation U_t increases above the regime threshold c^*, resuming the Stress regime. The highlighted period in the figure represents the Hurricane Ida event during August 2021, which demonstrates a real-life scenario where this strategy can be applied.

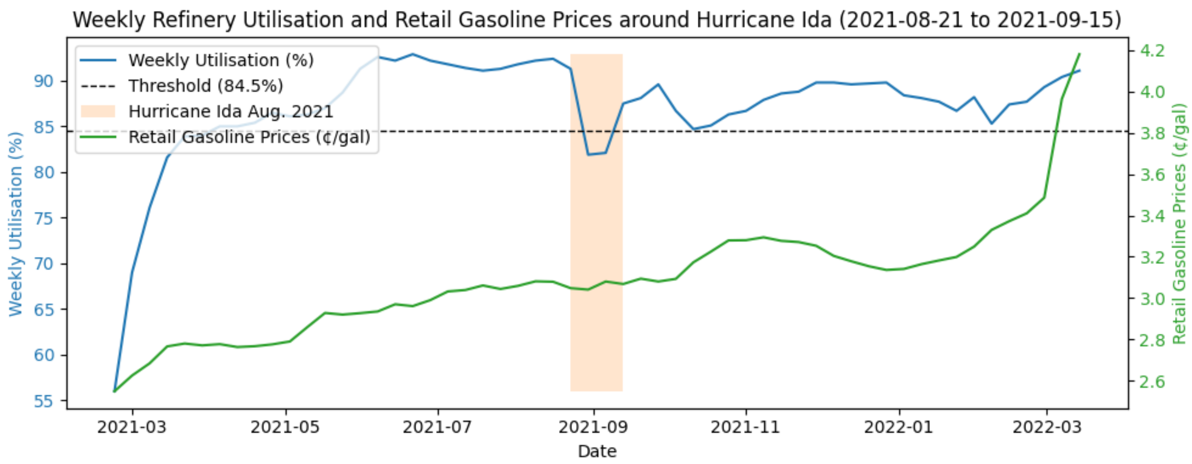

Figure 6 Hurricane Ida period time series of utilisation of retail gasoline price

Explanation of the strategy:

- As U_t falls below the threshold, the positive correlation between utilisation and price is expected. Here, we enter a long position expecting the price to rise as utilisation climbs back.
- As U_t climbs back up above the threshold, the negative correlation between U_t and $\Delta P_t^{(2)}$ is expected. Here, we exit the long position and transit to short position, expecting the price to fall as utilisation keeps increasing.

A schematic of the Regime Cross-Over Trading Strategy is illustrated below.

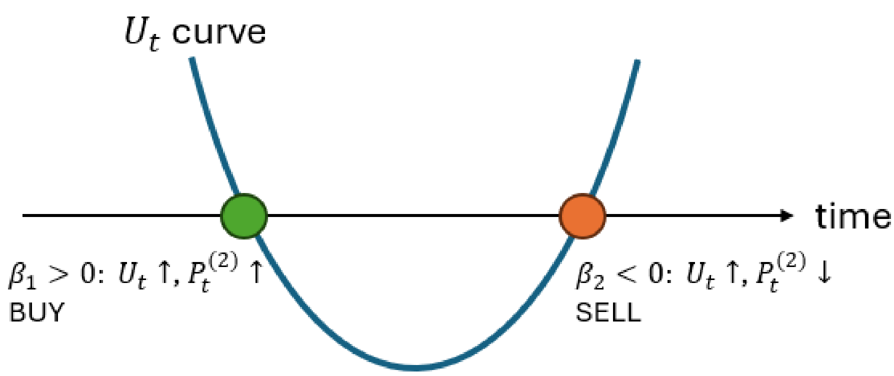

Figure 7 Schematic of proposed Regime Crossing Strategy

5.2 Backtesting

Out-of-sample back-tests are run to assess the effectiveness of this strategy during the period between 2025-01-01 to 2025-05-01, where such a utilisation dip is observed. The utilisation "dip" of interest is highlighted.

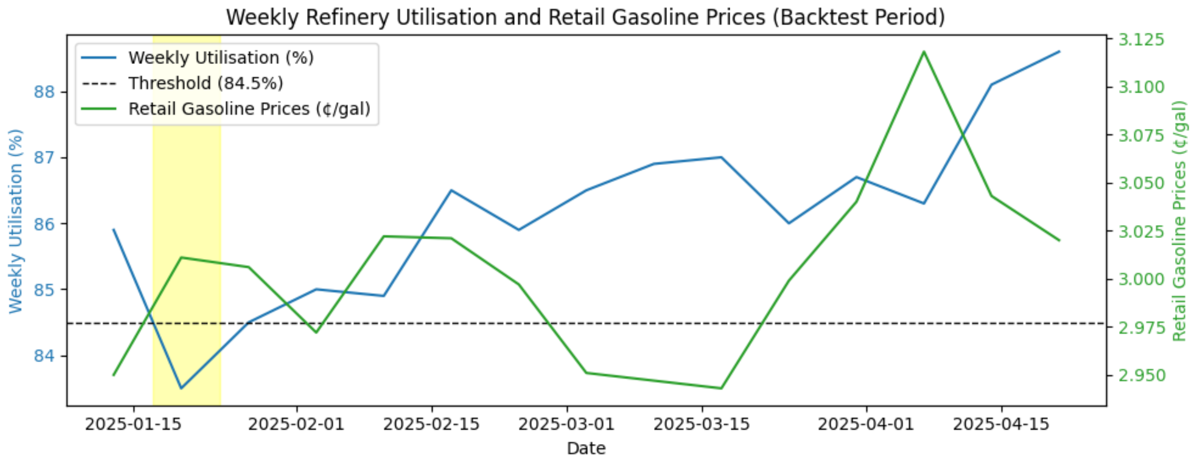

Figure 8 Backtest period time series

The forecast of price change is two weeks in advance; therefore, we enter a long position on 2025-01-27 and switch to a short position on 2025-02-10, as indicated by the highlighted period below. The cumulated PnL from the proposed regime crossing strategy is compared against the strategy using the simple single-slope predictor where the only trading signal is when the projected price change is positive.

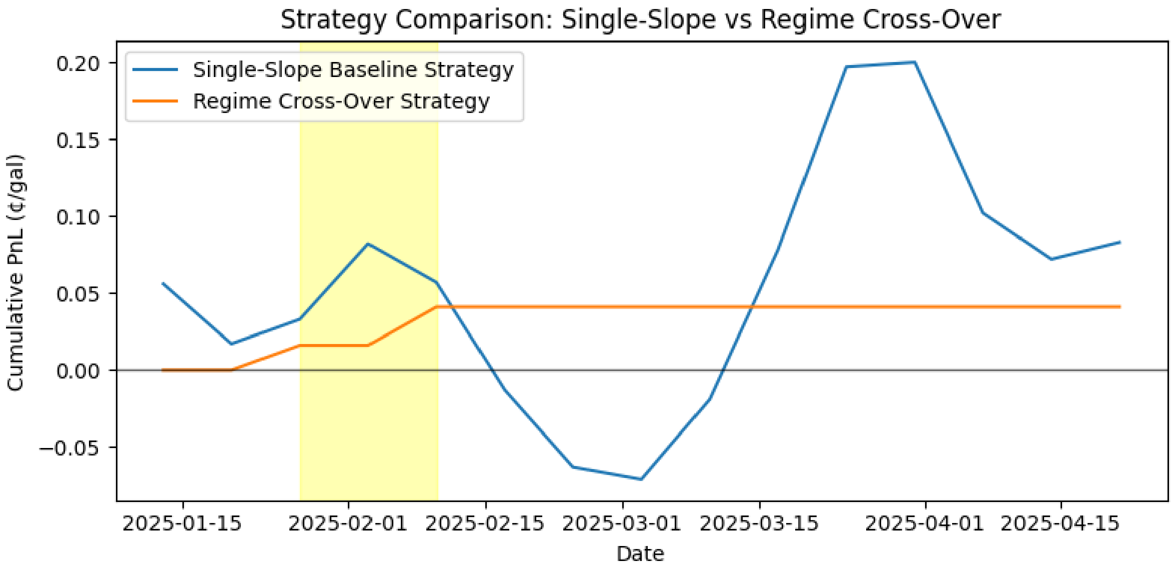

Figure 9 Backtest cumulative PnL of two strategies

Clearly, the newly proposed strategy prevents a draw-down which the single-slope predictor failed to capture. There is still much left to be implemented for the proposed strategy to profit from the regime-switching behaviour; however, we will simply illustrate the general idea here due to time constraints.

6. Conclusion

This study investigates and exploits a clear structural break in how U.S. refinery utilization impacts retail gasoline prices. A sup-LM test results establish 84.5% utilization as the threshold separating two distinct price-sensitivity regimes. In the "easy" regime ($\leqslant$ 84.5%), small positive price responses prevail; in the "stress" regime ($>$ 84.5%), price changes reverse sign and become more volatile. Rolling correlations and regime-wise OLS confirm this sign reversal and magnitude disparity.

Incorporating a simple binary Stress flag into a logistic spike-prediction model yields substantial gains—raising ROC-AUC from 0.69 to 0.74 and doubling Precision@100 from 3% to 7%. Our proposed "Buying the Dip" trading strategy, which capitalizes on transitions across the 84.5% threshold, demonstrates out-of-sample robustness by avoiding drawdowns that hampered a single-slope predictor.

These findings offer a practical early-warning metric for large gasoline price moves and lay the groundwork for dynamic trading or hedging strategies. Future research should refine this framework by incorporating transaction costs, optimizing position sizing, integrating macroeconomic indicators, and extending the analysis to intra-week frequencies and global markets to capture more nuanced regime shifts.

Reference

[1] *What Drives U.S. Gasoline Prices?* U.S. Energy Information Administration, October 2014