CFTC Positioning: Does It Improve Forecasts?

Can knowing what futures traders hold help predict energy prices and large price moves?

Data through: May 30, 2024
Model results: November 1, 2026
Write-up: November 4, 2026

1. Introduction

I wanted to find out whether knowing how traders are positioned helps predict what happens next.

The Commodity Futures Trading Commission (CFTC), the U.S. regulator for these markets, publishes weekly Commitments of Traders (COT) reports that count positions by trader group.

I mainly looked at managed money. I also used positions held by producers and other businesses handling the physical commodity, and by swap dealers. Swap dealers arrange contracts tied to commodity prices and may use futures to manage the risk from those contracts.

The questions I tested were:

Question	What would the extra information help predict?
Prices	Price changes one, four, or eight weeks ahead.
Risk	How much prices move around, or whether a large fall occurs, over four weeks.
Crowding	Whether a recent price rise or fall continues when many traders hold similar positions.
Price gaps	Changes in the gap between product and crude prices, or between two futures delivery months.

Question	What would the extra information help predict?
Inventory news	Price moves around unexpected increases or decreases in stored oil and products.

2. Data

Which markets and dates?

I used the CFTC reports that separate traders into business groups and count futures positions without options. The markets were West Texas Intermediate (WTI) crude oil, RBOB gasoline, and heating oil, now ultra-low-sulfur diesel (ULSD), traded on the New York Mercantile Exchange (NYMEX).

The U.S. Energy Information Administration (EIA) supplied prices and physical-market data, and inventories. The price targets were WTI, New York Harbor conventional gasoline, and New York Harbor ULSD spot prices.

The first training data start in 2014, after heating-oil futures changed to ULSD. I checked forecasts made from January 2018 to early April 2024. EIA’s daily futures series ends on April 5, 2024, so it limits the common comparison period even though other data continue later. The longest price outcomes end in May.

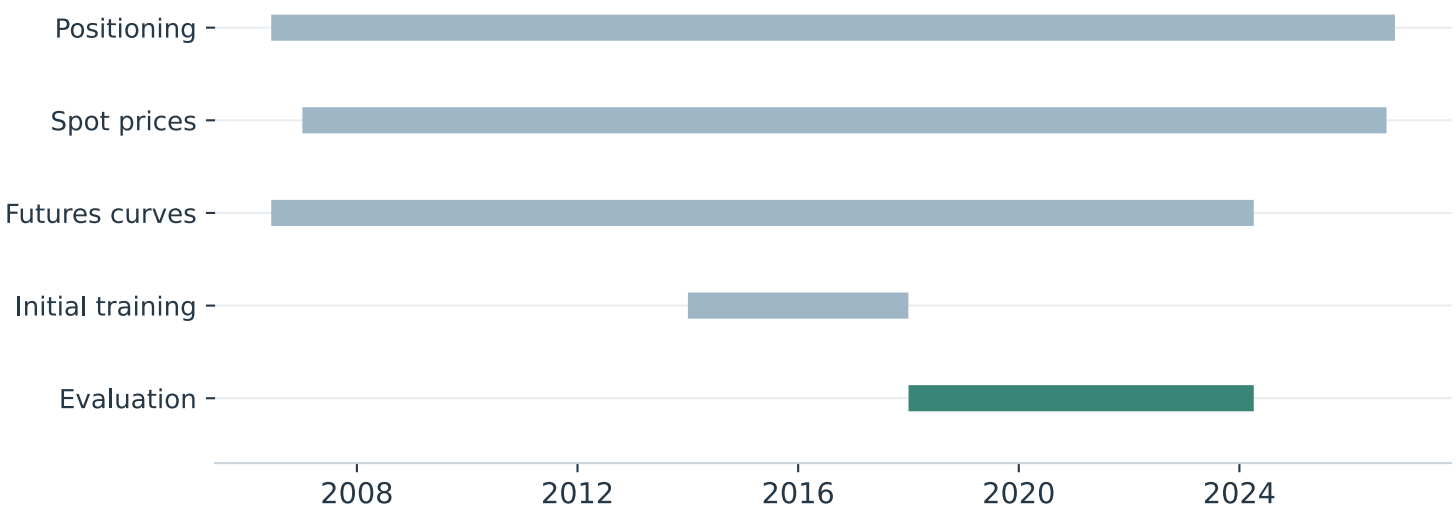

Figure 1. The dates available for each part of the study. CFTC reports continue into 2026, but EIA futures prices stop on April 5, 2024. I train the first models on data from 2014 and check forecasts from 2018 to April 2024, with their later outcomes through May.

How I measured positioning

Open interest is the number of futures contracts still open. I divide positions by open interest so that a larger market does not automatically look more heavily positioned.

Net exposure is long positions minus short positions, divided by open interest. Gross exposure adds long and short positions, excluding separately reported spread positions that pair positions in different contracts. I also used changes over one and four reports, and producer and swap-dealer net shares.

A percentile ranks the latest reading against the market’s past three years. A high percentile means that the reading is high compared with that history.

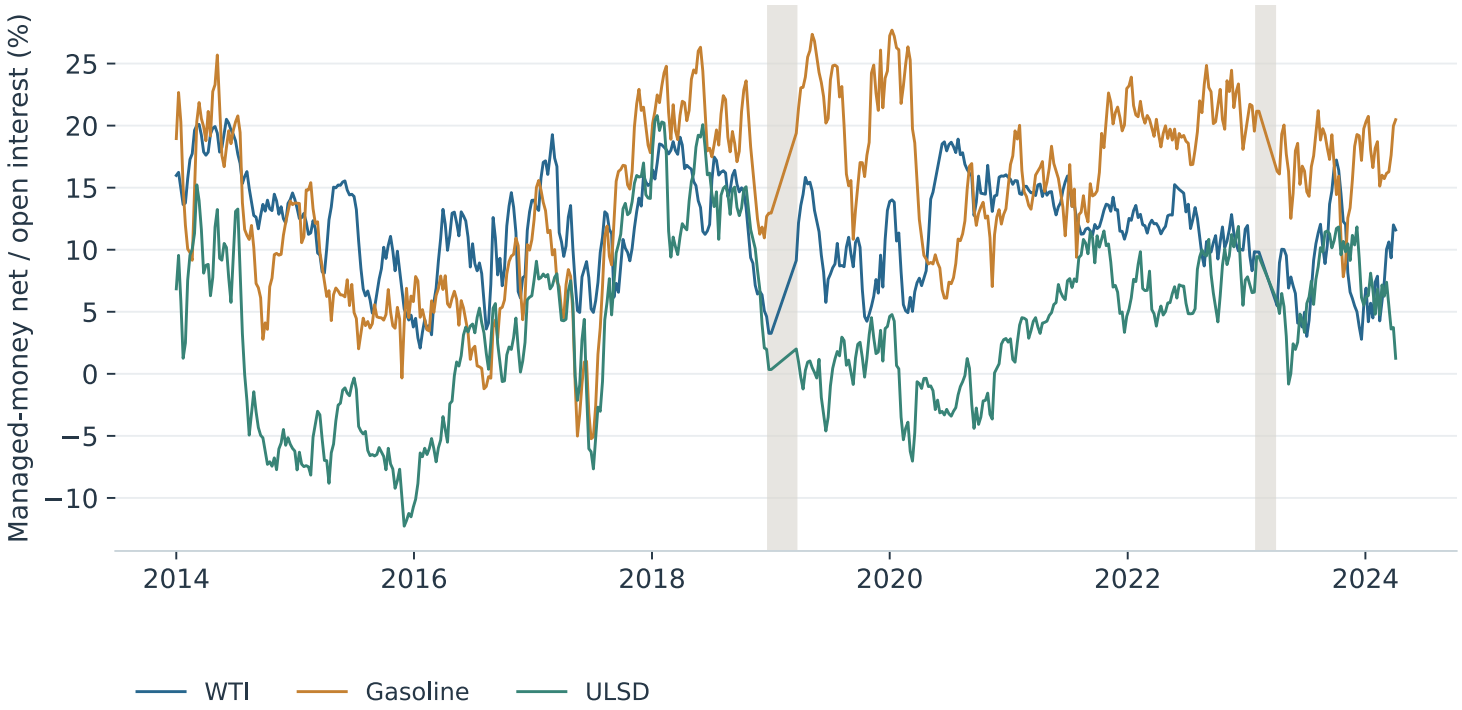

Figure 2. Funds' net futures positions. Long positions minus short positions, divided by all open contracts, gives the net share shown here. Shading marks delayed reports and the extra time excluded around them.

Could the information have been known in time?

CFTC usually measures positions on Tuesday and releases them on Friday. My forecasts use Wednesday decisions, so I wait eight calendar days after the position date before using a report.

Inventory inputs had to come from reports released before the decision, using the EIA report archive. Price inputs also had to be earlier than the decision. The future price change starts with the next available quote.

There are remaining limits. The histories contain later revisions, and I assume daily quotes become available with a one-day delay. I do not have the exact versions of the data first published at every date. Missing prices stay missing.

The models use log price changes. It is close to an ordinary percentage change for small moves, but can differ for large moves. This measure requires positive prices, so nonpositive prices, including the effects of negative WTI in April 2020, are left out. Most forecast comparisons have about 300 weekly observations per market; their exact counts differ.

3. Models

Compare forecasts with and without the extra data

The baseline model used recent price trends, past price variability, the price gap between nearby futures contracts, inventories compared with their usual seasonal level, the latest inventory change, and the time of year. Seasonal comparisons use earlier data only.

I then add positioning to the same model. The risk models also allow its effect to depend on inventories and the recent price trend. For the crowding hypothesis, both models already contain positioning. The added terms ask whether the effect of a recent price rise or fall changes when positions are unusually one-sided.

For outcomes measured as amounts, I use ridge regression. Regression estimates how inputs relate to an outcome; ridge adds a penalty that keeps its input weights—the amounts each input changes the forecast—from becoming too large. A related method, logistic regression, estimates the probability of a large price fall. Its weights are limited too. The amount models fit squared errors, which give large misses extra weight, but I mainly judge them by average absolute error.

I fix the settings and repeat the work with weaker and stronger penalties and with the combined CFTC report. I refit every quarter, adding earlier observations to the training history. Every training outcome must already have finished. For example, a forecast's eight-week result cannot enter training while those eight weeks are still passing. Input limits and scaling are also calculated from the training data alone.

Metrics used to measure models

Mean absolute error (MAE) is the average size of a miss, ignoring its direction. Lower is better. Price errors are in percentage points of log price change; price-gap errors are in dollars per barrel. Errors in price variability are expressed on a one-year scale, even though the forecast covers four weeks.

For probabilities, I use the Brier score. For each forecast, I take the gap between its probability and the outcome—one if the event happened, zero if it did not—square that gap, and average across forecasts. A lower score is better.

> Error reduction = 100 × (baseline error − error with added information) ÷ baseline error.
> Above zero means smaller errors. Below zero means larger errors.

Longer forecasts overlap as two forecasts made a week apart include many of the same future days. Their errors are therefore related. To estimate uncertainty, I resample paired errors 1,000 times in blocks of twelve observations, keeping some of that relationship. It produces the 95% uncertainty ranges in the tables and charts. A wider range means a less precise estimate. If it includes zero, the test does not show a clear improvement.

Trying many comparisons makes it easier to find a small gain by chance. I adjust the statistical checks across all 26 main tests. A gain counts as supported only if its average is positive, its uncertainty range stays above zero, and its adjusted p-value is below 0.05.

4. Results

Price forecasts

The simple no-change forecast wins. Positioning slightly lowers gasoline error at all three forecast lengths. It raises WTI error at all three and has mixed effects on ULSD. None of the gains passes the statistical checks.

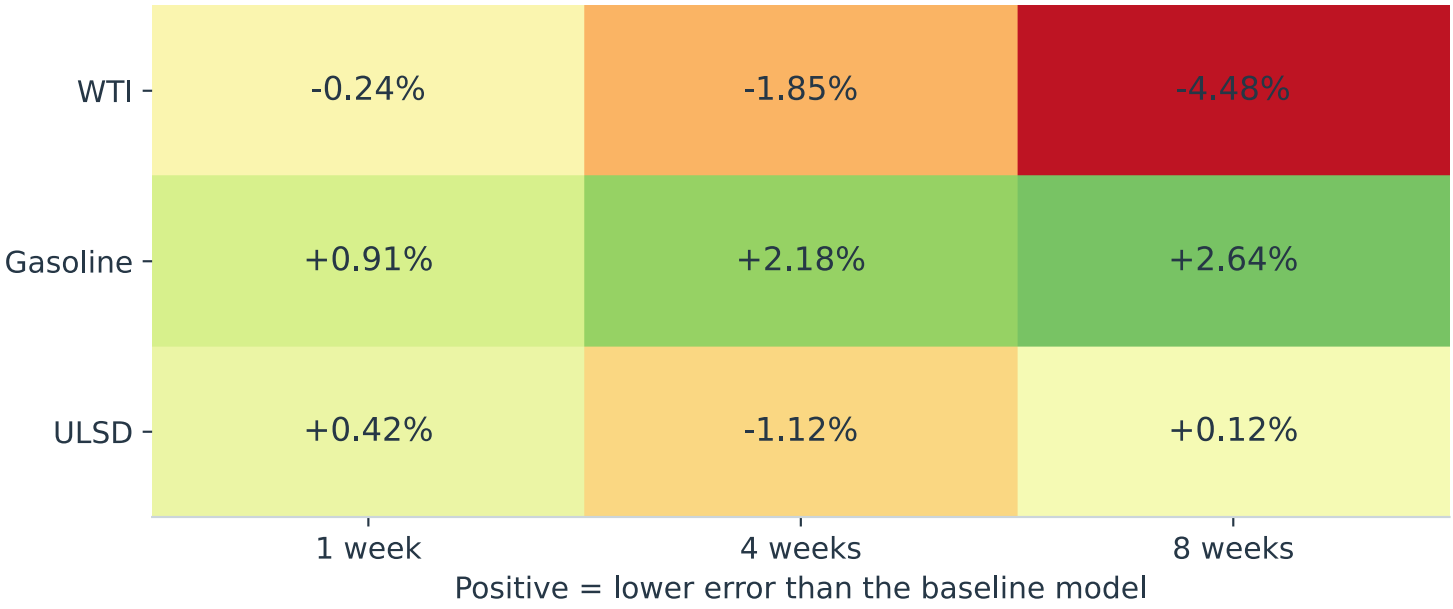

Figure 3. Prices: small and mixed changes in price error. Each value shows how much average absolute error (MAE) changes after adding positioning. Gasoline improves slightly at each forecast length; WTI worsens.

Market and forecast	Starting model error	Error with CFTC data	Error reduction	95% uncertainty range
WTI · 1-week price change	4.12	4.13	-0.24%	-2.00% to +1.40%
WTI · 4-week price change	8.65	8.81	-1.85%	-4.51% to +1.03%
WTI · 8-week price change	13.27	13.86	-4.48%	-7.76% to -1.04%
Gasoline · 1-week price change	4.49	4.45	+0.91%	-1.52% to +3.33%
Gasoline · 4-week price change	9.70	9.49	+2.18%	-3.77% to +8.93%
Gasoline · 8-week price change	14.86	14.46	+2.64%	-4.75% to +9.45%
ULSD · 1-week price change	4.26	4.25	+0.42%	-1.54% to +2.84%

Market and forecast	Starting model error	Error with CFTC data	Error reduction	95% uncertainty range
ULSD · 4-week price change	9.36	9.46	-1.12%	-8.16% to +5.34%
ULSD · 8-week price change	14.50	14.48	+0.12%	-9.80% to +9.46%

For four-week gasoline forecasts, MAE falls from 9.70 to 9.49 percentage points, a 2.18% reduction. But predicting no change has an even smaller error of 9.36 points. That simple rule beats both fitted models in all nine price comparisons.

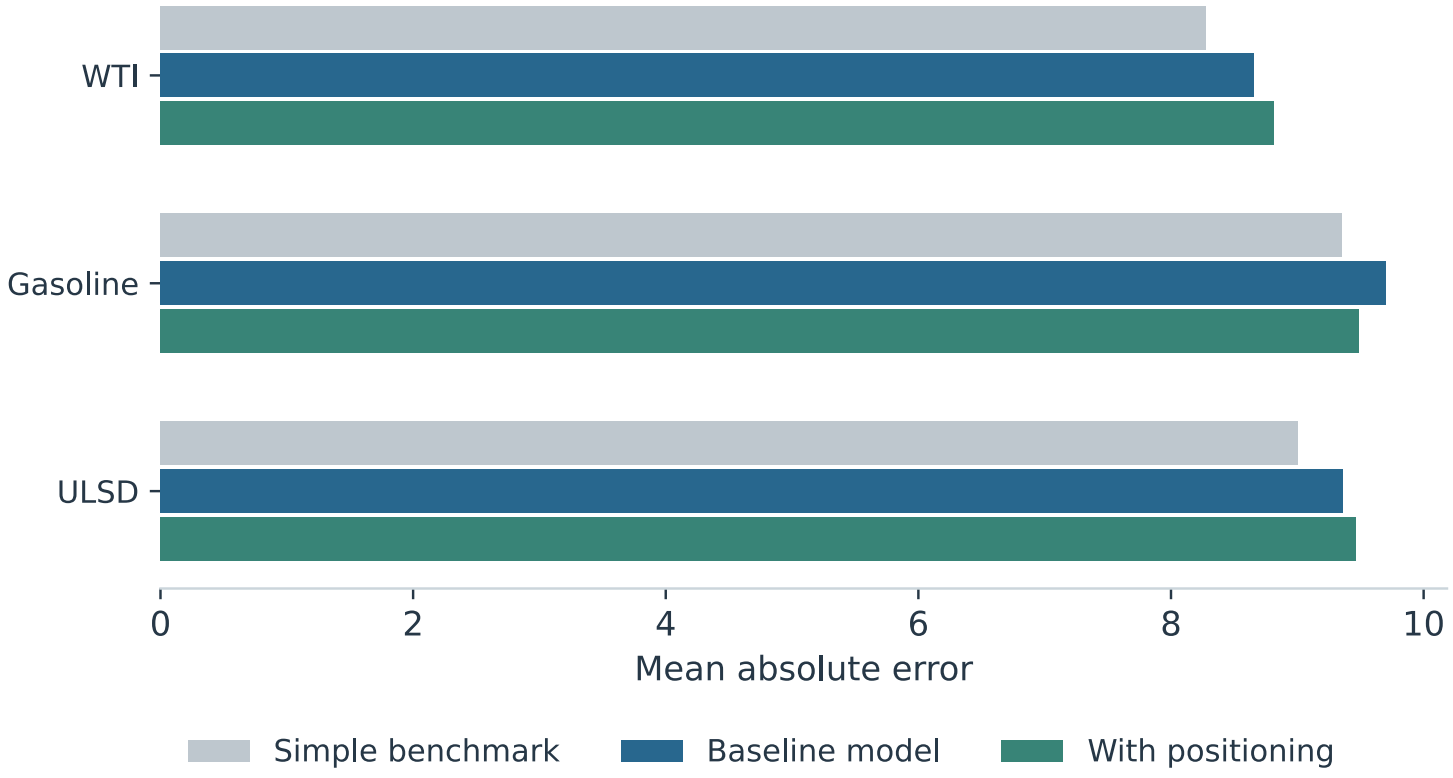

Figure 4. Predicting no change beats both price models. For four-week forecasts, predicting no price change has lower average error than either fitted model in all three markets. The unit is percentage points of log price change.

The small gasoline gain is also uneven over time. Its largest four-week improvement is in 2021; several other years get worse. The overall average hides these differences.

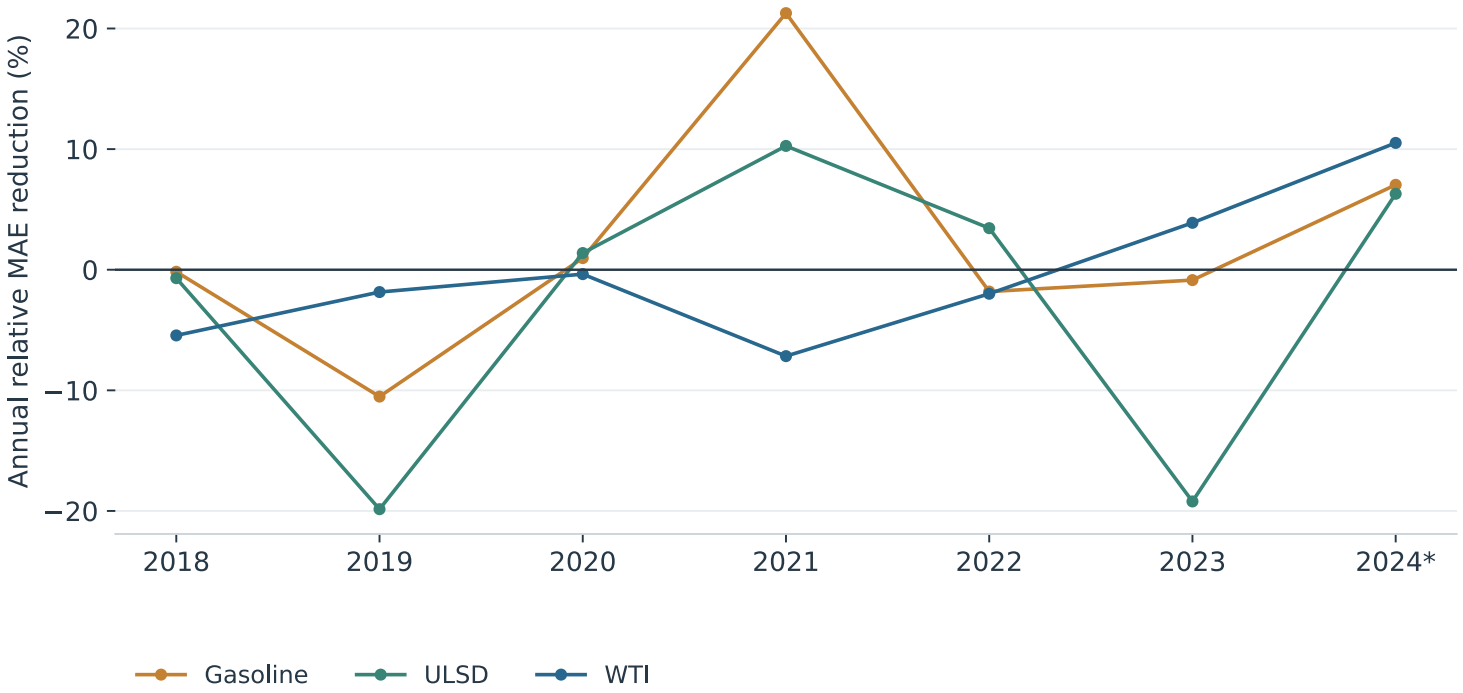

Figure 5. The price results change from year to year. The chart separates four-week error reductions by forecast year. Gasoline's largest gain is in 2021, while several years get worse. The partial 2024 sample has only fourteen decisions.

Price variability and large falls

Most risk forecasts get worse. Volatility means how much prices move around. Here I measure the variability of daily log price changes over the next four weeks, then put it on a one-year scale. A large-fall event means a quote at least 10% below the starting price within four weeks.

WTI volatility MAE rises 21.8%. ULSD error falls by less than 1%, but the uncertainty is large. The probability forecasts for large falls get worse in all three markets.

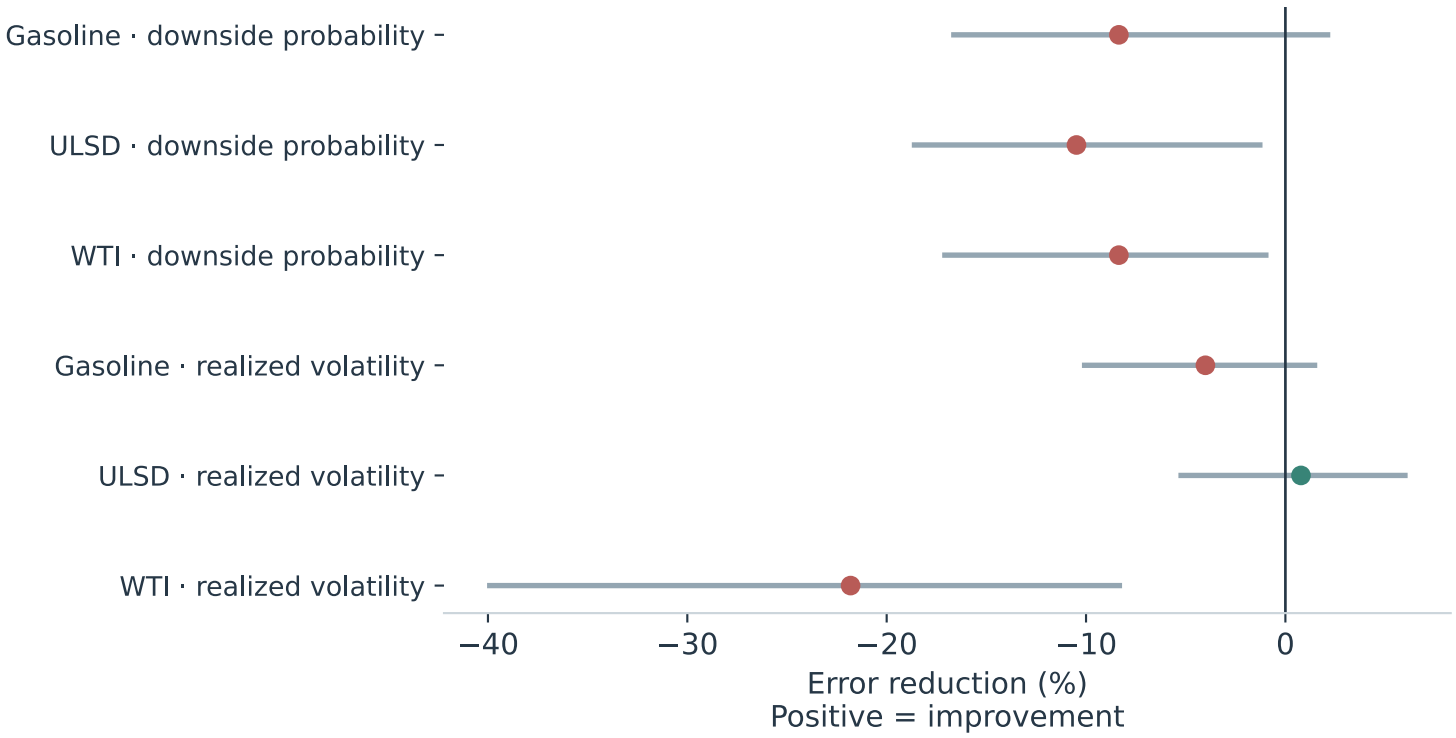

Figure 6. Risk: most risk forecasts get worse. Dots show the percentage reduction in forecast error; lines show 95% uncertainty ranges from resampling related forecasts together. Price variability uses MAE on a one-year scale; large-fall probabilities use Brier score. These percentages describe changes in error, not changes in the probability of a fall.

Market and forecast	Starting model error	Error with CFTC data	Error reduction	95% uncertainty range
WTI · Observed price variability over four weeks	13.88	16.91	-21.81%	-39.91% to -8.33%
WTI · Chance of a large price fall within four weeks	0.1728	0.1872	-8.35%	-17.08% to -0.98%
Gasoline · Observed price variability over four weeks	14.46	15.04	-4.01%	-10.07% to +1.46%
Gasoline · Chance of a large price fall within four weeks	0.1983	0.2149	-8.35%	-16.63% to +2.12%
ULSD · Observed price variability over four weeks	12.91	12.81	+0.78%	-5.23% to +6.00%

Market and forecast	Starting model error	Error with CFTC data	Error reduction	95% uncertainty range
ULSD · Chance of a large price fall within four weeks	0.1856	0.2050	-10.48%	-18.61% to -1.28%

The next chart asks whether events happened about as often as their predicted probabilities suggested.

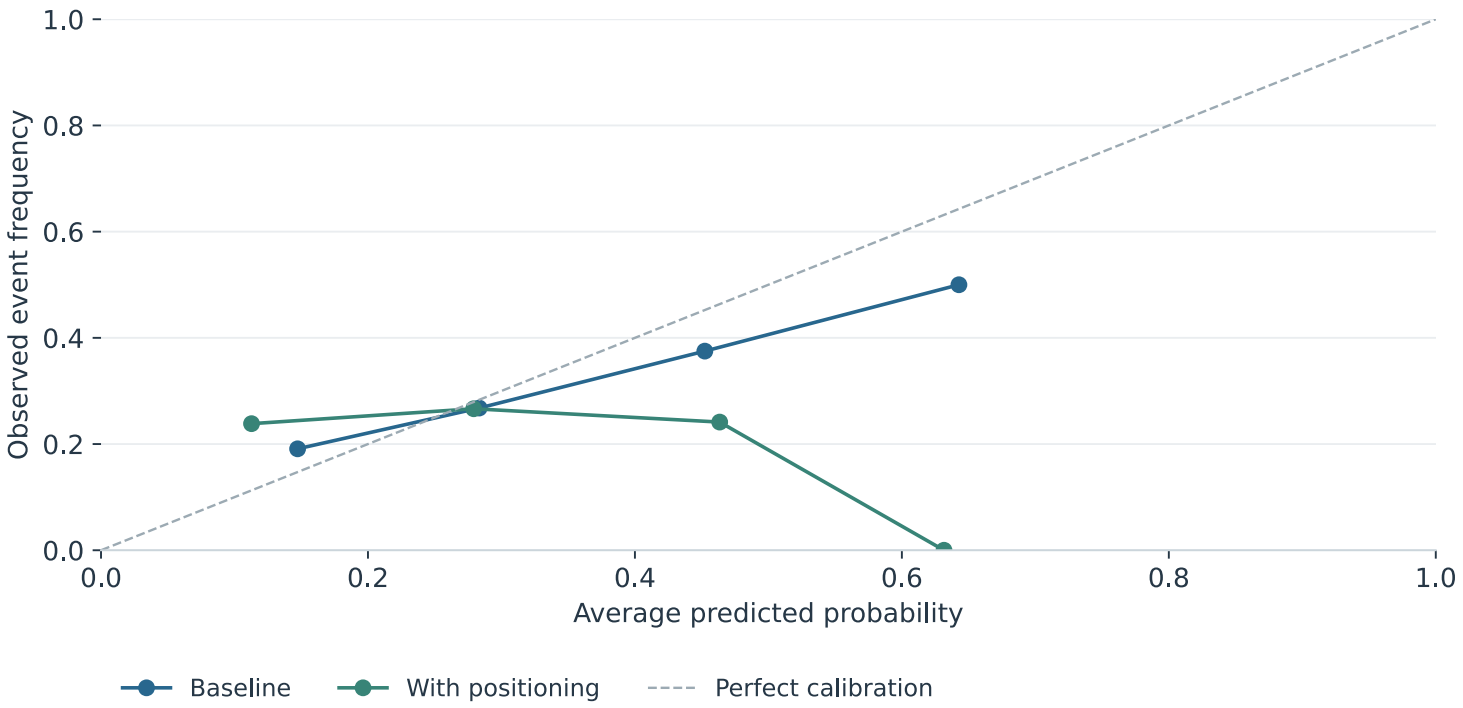

Figure 7. Predicted chances compared with what happened. A large-fall event means a price at least 10% below its starting level within four weeks. Forecasts with similar probabilities are grouped across all three markets. Points on the dashed line would mean that predicted chances match observed frequencies.

One-sided positions and recent price trends

The striking historical examples do not become a reliable forecasting rule**.** I call positions “crowded long” at or above their past 80th percentile and “crowded short” at or below the 20th. The rest are the middle group. I also split cases by whether prices rose or fell over the previous four weeks. That recent direction is called momentum.

After a prior price rise, gasoline's average next four-week log-price change is about +11.7% in the crowded-short group and −6.9% in the crowded-long group. But the first group has only nineteen observations, outcomes overlap, and other markets behave differently. WTI has just eight crowded-long cases following a price fall.

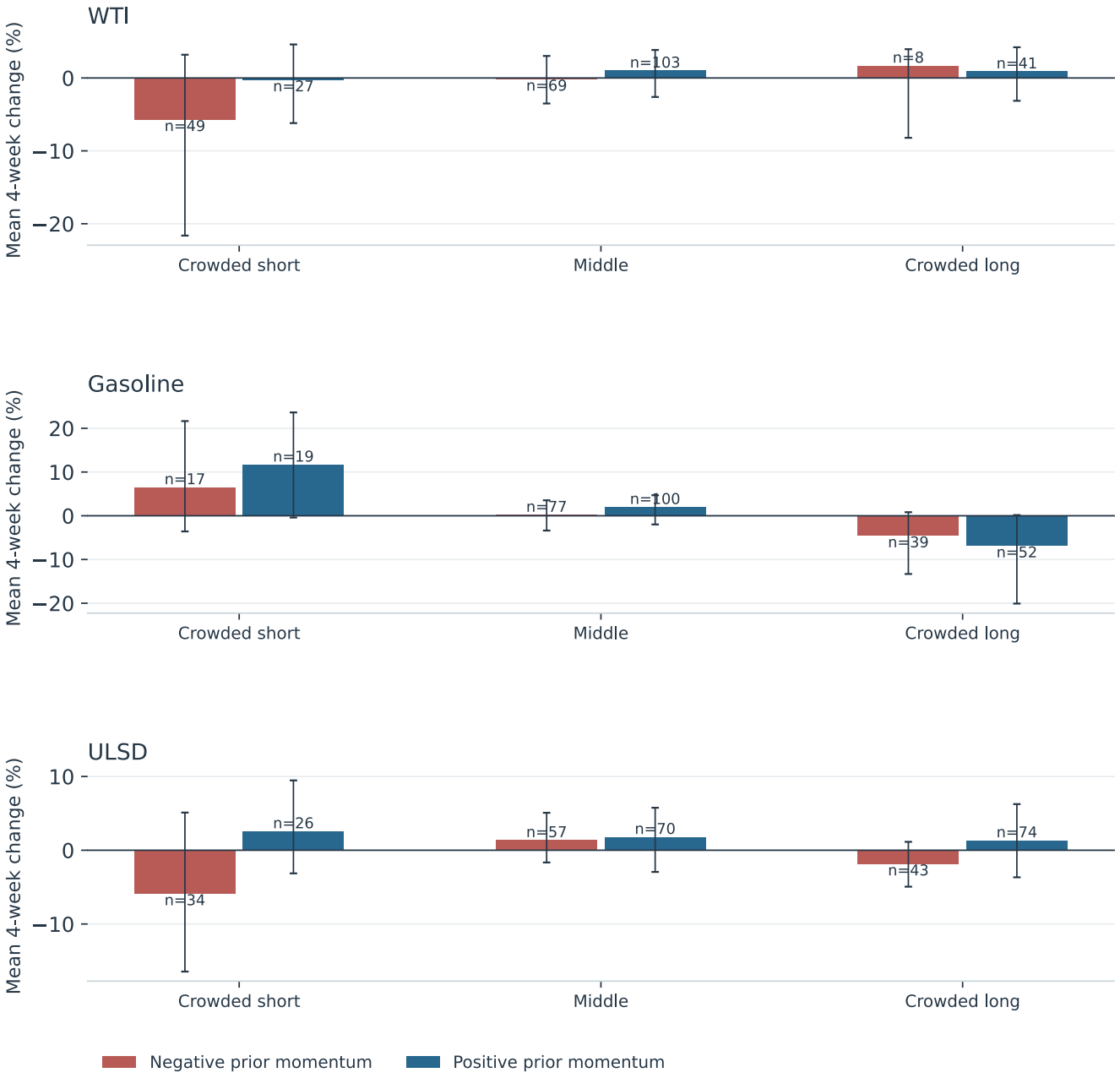

Figure 8. Crowding: one-sided positions and later price changes. The bars show average four-week log-price changes, grouped by the prior price trend and positioning. "Crowded long" and "crowded short" are the top and bottom 20% of past positioning. Counts and approximate 95% uncertainty ranges matter: some groups have very few observations.

Letting momentum's effect depend on crowding lowers gasoline error by only 0.71%. WTI and ULSD worsen slightly. Every uncertainty range includes no improvement.

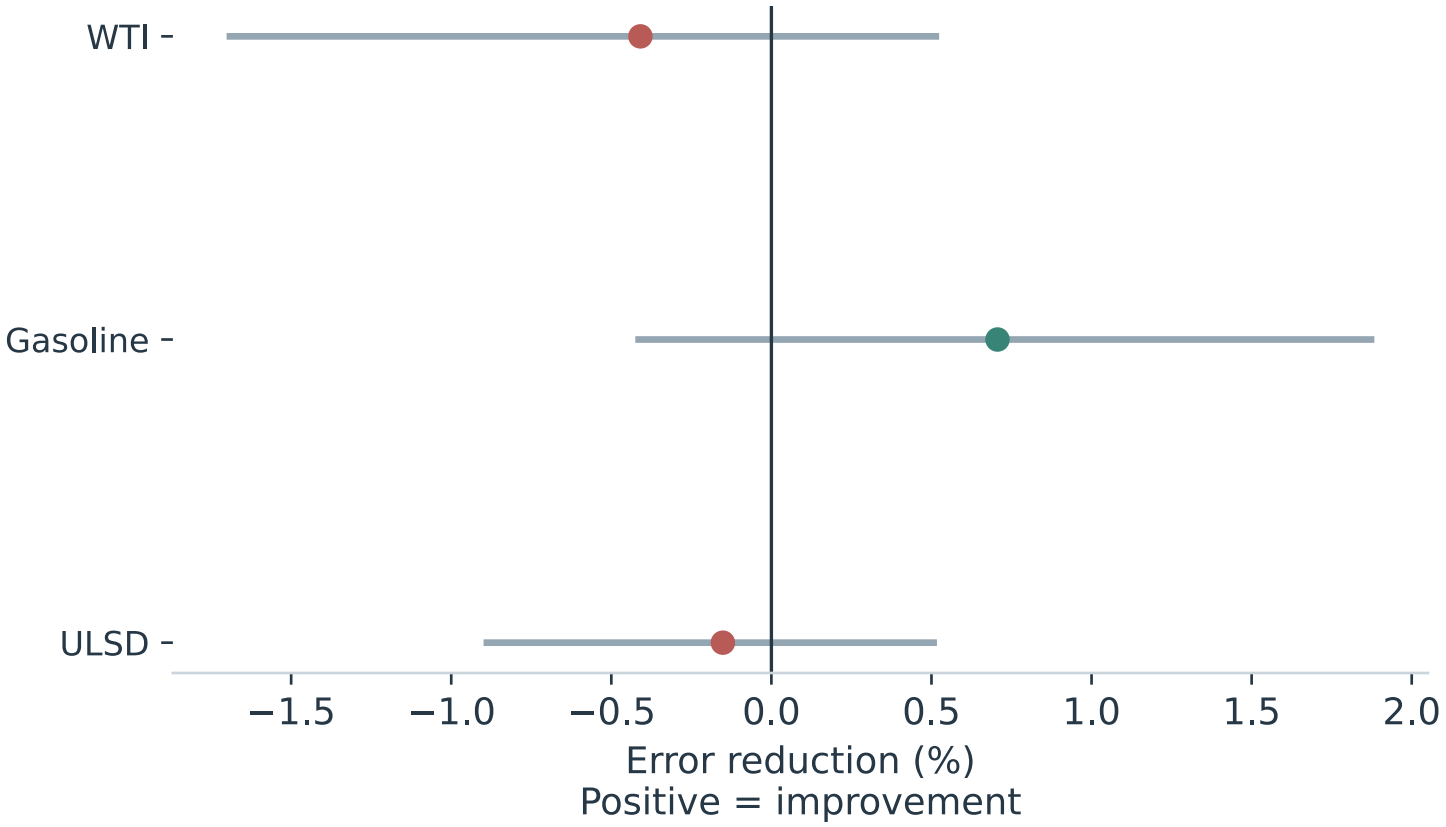

Figure 9. Adding crowding terms does little for forecasts. Both starting models already use positioning. The added terms let the effect of the recent price trend depend on how one-sided positions are. Gasoline's 0.71% error reduction is small and uncertain. No market has a supported gain.

Market and forecast	Starting model (with CFTC)	With added crowding terms	Error reduction	95% uncertainty range
WTI	8.81	8.85	-0.41%	-1.69% to +0.51%
Gasoline	9.49	9.42	+0.71%	-0.41% to +1.87%
ULSD	9.46	9.48	-0.15%	-0.89% to +0.51%

Gaps between prices

Positioning adds little to forecasts of price gaps. A **spread** is the difference between two prices. Here a **crack reference** is gasoline or diesel spot price minus WTI crude price. I first convert product prices into dollars per barrel using 42 gallons per barrel. A **calendar reference** is the futures price for the nearest delivery month minus the next month's price. I forecast how these gaps change over four weeks.

The product-minus-crude models add the difference between product and crude positioning. The delivery-month models add positioning for that market. Gasoline's calendar reference improves by just 0.22%; the other four comparisons worsen. None has a supported gain.

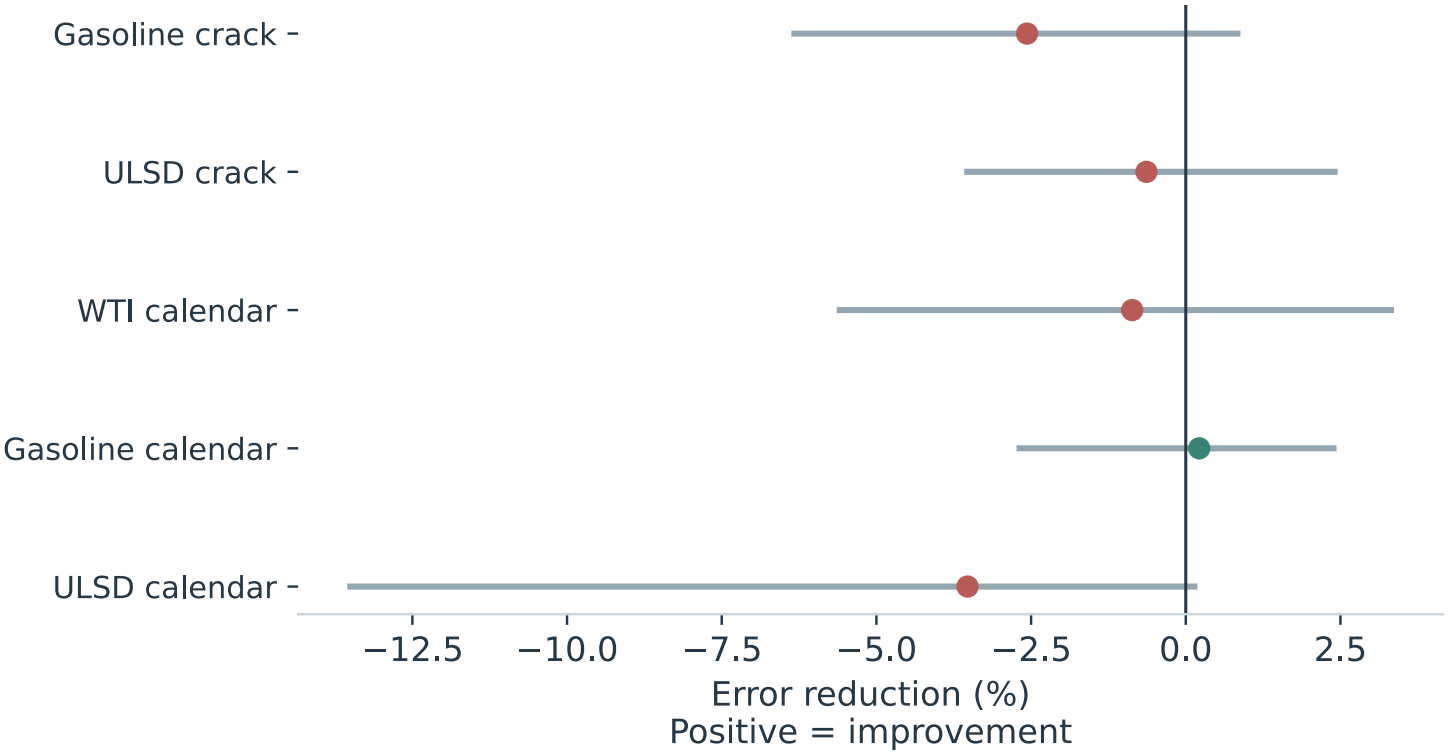

Figure 10. Price gaps: little help in predicting price gaps. I predict four-week changes in two product-minus-crude price gaps and three delivery-month price gaps, measured in dollars per barrel. Product prices use 42 gallons per barrel for conversion. Gasoline's delivery-month gap is the only improvement, at 0.22%, and it fails the statistical checks.

Market and forecast	Starting model error	Error with CFTC data	Error reduction	95% uncertainty range
Gasoline physical crack reference	4.38	4.50	-2.57%	-6.33% to +0.83%
ULSD physical crack reference	6.41	6.45	-0.64%	-3.53% to +2.41%
WTI nearby calendar reference	0.45	0.45	-0.87%	-5.59% to +3.32%
Gasoline nearby calendar reference	2.07	2.07	+0.22%	-2.69% to +2.39%

Market and forecast	Starting model error	Error with CFTC data	Error reduction	95% uncertainty range
ULSD nearby calendar reference	1.33	1.38	-3.53%	-13.50% to +0.14%

Some starting models beat the no-change rule, but adding CFTC data does little to help them.

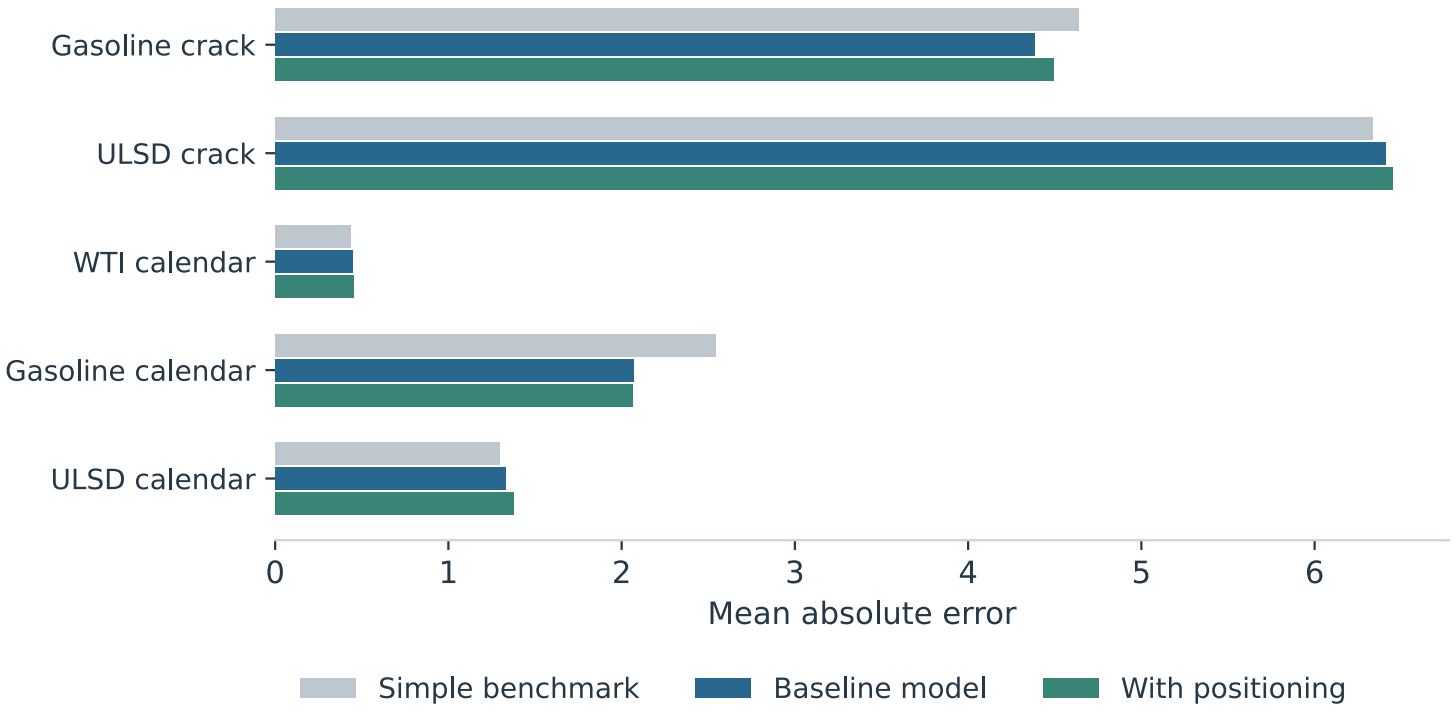

Figure 11. Price-gap forecasts compared with no change. Bars show MAE in dollars per barrel. Some starting models beat a forecast of an unchanged price gap, but positioning adds little. Changing delivery contracts and differences in product grade and location mean that these results do not show actual trading profits.

Inventory news

The daily test finds no gain, but the reaction to news remains a question. An inventory build means more barrels in storage; a draw means fewer. To test an unexpected report properly, I would need the market’s expectation before publication, the figures originally released, and prices measured just before and after the announcement. I do not have access to all of that information.

I therefore estimate expected weekly stock changes from earlier reports, recent changes, and seasonal patterns. The observed change minus that prediction is my surprise proxy. I scale it by past variability so that unusual changes stand out.

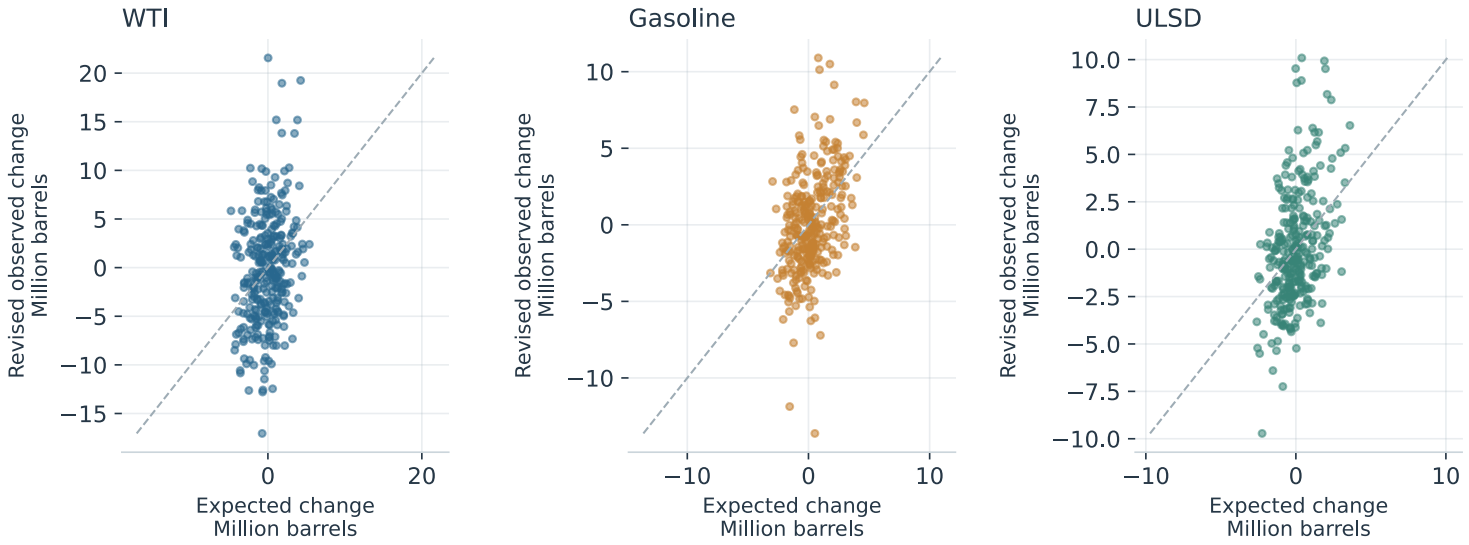

Figure 12. Inventory news: estimating what would count as an inventory surprise. I predict weekly stock changes using earlier reports and seasonal patterns. The difference between the later revised figure and that prediction, scaled by past variability, is the surprise proxy.

The starting model relates the whole release day’s spot-price change to all three estimated surprises, earlier prices, price variability, and seasonality. I then add positioning and a term that lets the selected market’s surprise matter differently depending on positioning. I leave out dates with multiple weekly releases.

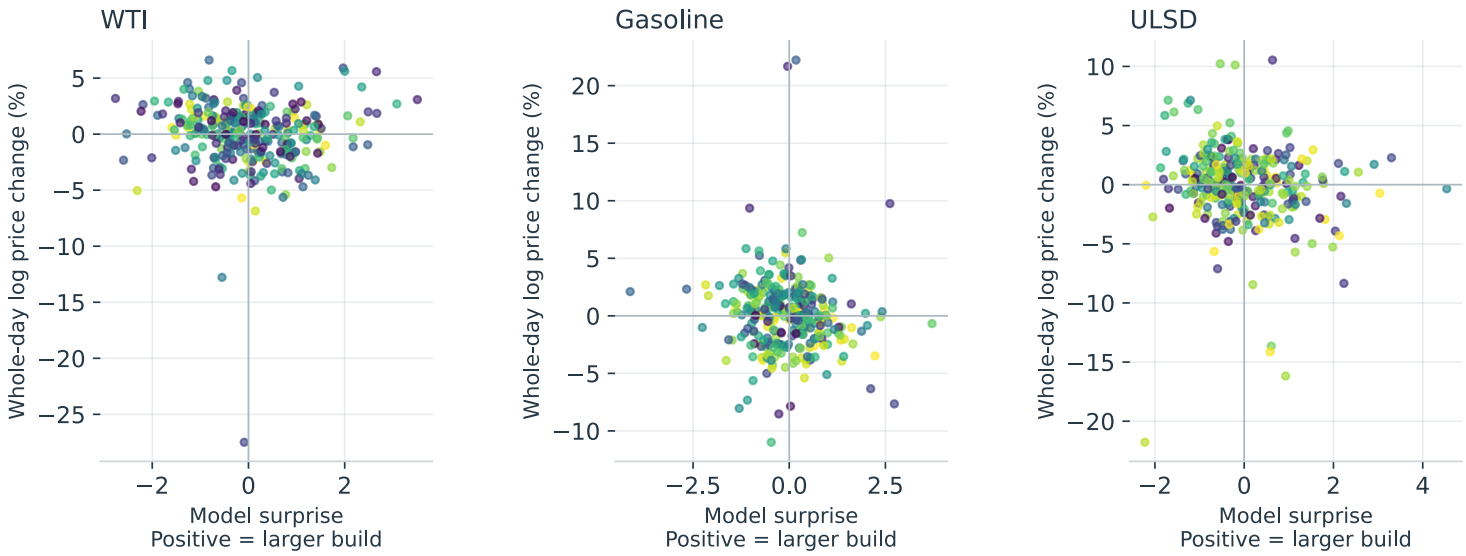

Figure 13. Estimated surprises and prices on the release day. Each point compares an estimated inventory surprise with the whole day's spot-price change. Positive surprises mean a larger stock build, or a smaller draw, than the model expected. Color shows positioning from low (purple) to high (yellow). Other news also affects daily prices, so this chart cannot isolate the announcement's immediate effect.

Adding positioning slightly raises MAE in all three markets. A whole day's price move includes other news. The surprise is also known only when the report arrives, so this test is neither a forecast made before the report nor proof that the announcement caused the move.

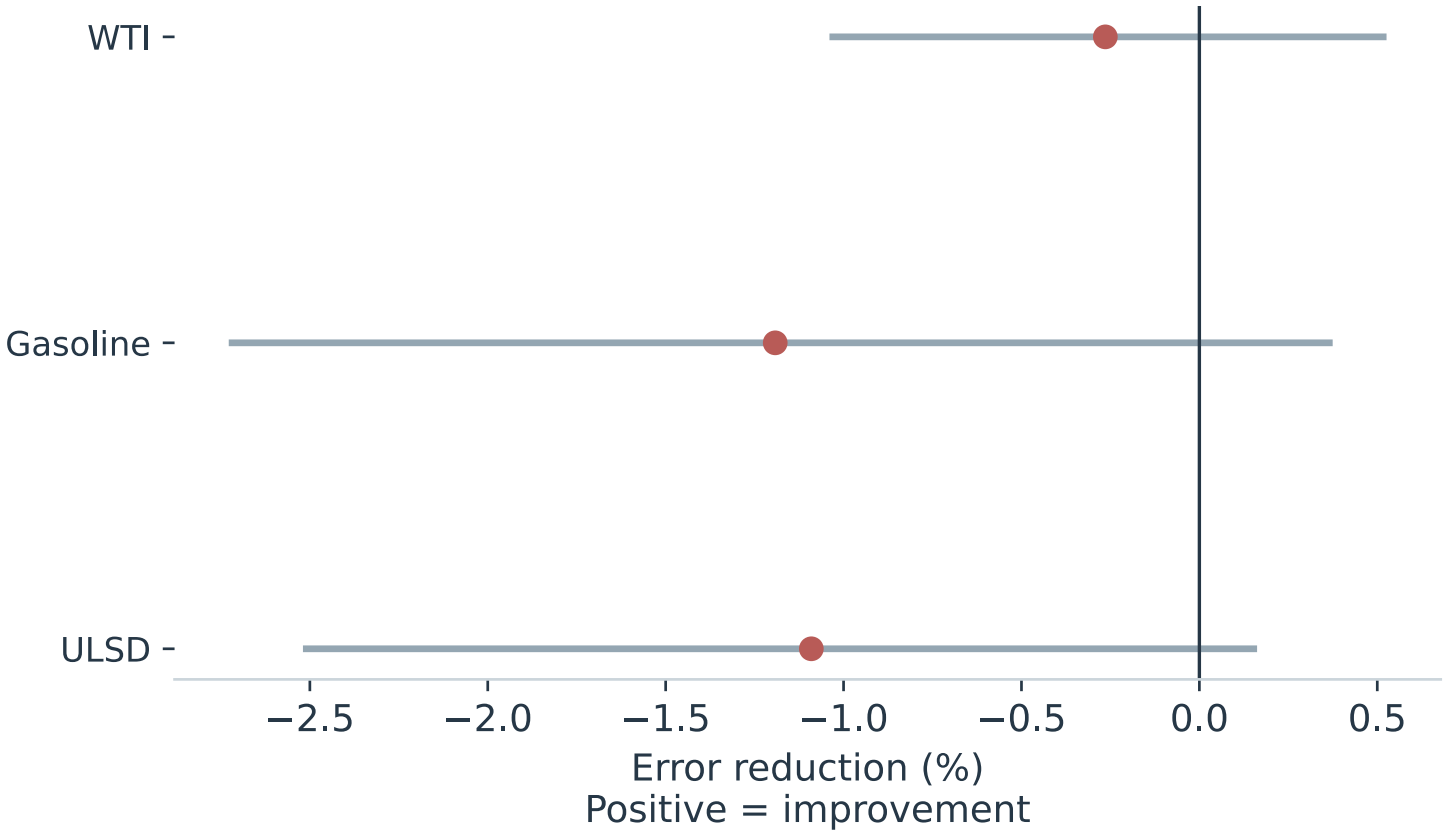

Figure 14. Positioning does not improve the whole-day test. The starting model uses the estimated inventory surprises, earlier prices, price variability, and seasonality. Letting the effect of a surprise depend on positioning slightly raises MAE in all three markets. The missing analyst expectations and intraday prices leave the immediate news-reaction question open.

Market and forecast	Starting model error	Error with CFTC data	Error reduction	95% uncertainty range
WTI	2.00	2.00	-0.26%	-1.03% to +0.52%
Gasoline	2.19	2.22	-1.19%	-2.72% to +0.37%
ULSD	2.10	2.12	-1.09%	-2.51% to +0.15%

A stronger test needs time-stamped analyst expectations, original releases, and prices from a short period around each announcement.

Do the model choices change the answer?

Weaker and stronger penalties and the combined futures-and-options report change the size of the four-week price gains. Gasoline stays slightly positive. ULSD switches from worsening to a small improvement with the stronger penalty.



Figure 15. Changing the settings gives no strong evidence. These are four-week price MAE reductions with weaker or stronger penalties on the model's weights, and with the separate report that includes options.

Across all 26 main comparisons, 8 lower average errors still lead to 0 supported improvements. This finding applies to the models, dates, and timing rules tested here.

5. Conclusion

I did not find reliable evidence that adding CFTC positioning improved these forecasts. Of the 26 main comparisons, 8 had lower average error, but none passed all the checks for chance results.

- **Prices:** the small gasoline gains still lost to predicting no price change.
- **Risk:** most forecasts of price variability and large falls got worse.
- **Crowding and price gaps:** the extra information added little dependable forecasting value.
- **Inventory news:** the daily test showed no gain, but could not settle what happens immediately around an unexpected announcement.

Limits and the data that would make the study more effective

The clearest gaps are in the prices and information available at each decision time. More detailed data would let me test actual futures contracts, measure inventory surprises against market expectations, and check reactions around an announcement. Those are proposed extensions; they do not change the results reported above.

Historical futures prices, trades, and quotes

I would first investigate CME DataMine for individual WTI, RBOB gasoline, and ULSD contracts, including expired contracts and observations after April 2024. A consistent contract history would extend the short comparison period and replace the spot-price targets with futures outcomes that better match the markets in the CFTC reports. I would retain delivery months and define in advance when a position moves to the next contract. A continuous chart alone would not be enough: its construction can hide contract changes and alter measured returns.

For the inventory-news test, I would add time-stamped trades and bid and ask quotes around each release. CME documents historical top-of-book data containing the best bids and offers and their sizes. These would let me measure moves over minutes instead of the whole day, and estimate the cost of crossing the bid–ask spread. Detailed order-book depth would be a later purchase if trade size became part of the question. Prices alone would still leave commissions, execution delays, and the rules for entering and exiting a trade to be specified.

Analyst expectations recorded before inventory reports

My inventory surprise is measured against my own model. The more useful question is whether prices react differently when EIA reports something the market did not expect. LSEG’s Reuters Polls is a candidate to investigate: LSEG documents historical consensus estimates and individual contributor forecasts, including commodity indicators. I would request the specific weekly U.S. crude, gasoline, and distillate inventory survey histories, their publication times, and any updates before each EIA announcement. General access to polling data would not by itself establish that this exact backfile is included.

With suitable coverage, I could calculate the first reported stock change minus the last consensus available before release. Individual forecasts would also show how much analysts disagreed. I could then test whether positioning changes the response to the surprise while accounting for simultaneous crude and product news. That would be a stronger test than treating an error in my inventory model as the market's surprise.

The earlier API inventory report

The American Petroleum Institute's Weekly Statistical Bulletin is a separate paid source available through its authorized distributors, LSEG and ICE. It normally arrives on Tuesday afternoon, before the EIA report, with holiday exceptions. Its historical crude, gasoline, and distillate estimates could help capture information traders already had before EIA published. I would need the original reports and actual release times, rather than a later summary. API and EIA are different surveys, so I would use API as earlier information, not as a replacement for the EIA outcome.

Data as first published, with release and revision times

Downloading today's history can give a past forecast information that was revised later. Macrobond's revision-history tools are one paid option to investigate for retrieving earlier versions of time series. Its documentation explicitly says that coverage depends on the series and requested period. I would verify the exact EIA stock series, first-release values, and release-time metadata in a sample before relying on it. Archived EIA reports could also recover some original figures without buying another service. For both inventories and CFTC reports, I would keep the original publication time, subsequent corrections, and the time each observation became usable.

Refinery activity and other physical-market information

S&P Global Energy's fuels and refining services are candidates for refinery operations, outages, trade flows, and supply-and-demand information. These could improve the starting model by distinguishing a crowded market from a market responding to a refinery disruption or a changing product balance. I would need historical observations and the dates they became available, not just today's reconstruction. This is a second priority after prices, expectations, and release timing: more explanatory variables cannot repair a mismeasured announcement surprise.

What better data would still leave unresolved

CFTC positioning remains a weekly snapshot of broad trader groups across delivery months. More detailed price data would not reveal those traders' individual motives or their positions immediately before an announcement. The current sample is also small for unusual crowding episodes, and overlapping forecasts do not become independent simply because I add more intraday observations. I would keep the existing results as a record, set the new tests before examining their outcomes, and reserve later periods for evaluation. Where a stronger starting model uses paid fundamentals, I would compare it with and without positioning on the same observations.

My spending priority would be a sample of contract-level futures prices and quotes, together with a verified inventory-consensus backfile; then historical API releases and any missing original-release records; then refinery and other physical-market data. Before committing, I would check instrument and date coverage, timestamps, revision handling, and permission to export data for reproducible research. The aim would be a more credible test of whether positioning adds information. Better data could confirm the weak results as well as reveal an improvement.

Data and code

Sources: CFTC historical reports and definitions; EIA spot prices, futures prices, and weekly inventory archive.

Download data and code · Download results · Study manifest.