MACHINE LEARNING IN COMMODITY FUTURES: BRIDGING DATA, THEORY, AND RETURN PREDICTABILITY

Tony Guida
Senior Quantitative Analyst, R&D, Atonra

During the past decade, machine learning (ML) methods have become foundational in quantitative equity investing, where datasets spanning traditional fundamentals to alternative signals have enabled the modeling of nonlinear relationships and cross-sectional return drivers. In contrast, commodity futures remain an underexplored frontier for ML applications, despite the growing maturity of commodities as a financialized asset class and their rising prominence in institutional portfolios.

This asymmetry arises from multiple structural frictions. First, commodities are not capital assets in the traditional sense. They are tangible goods, often classified as "consumption assets" (Lucas 1978), and lack a balance sheet or accounting disclosure framework comparable to that of stocks. Their value is shaped by physical supply chains, geopolitical disruptions, seasonality, and idiosyncratic features, such as storage costs or delivery mechanisms—factors that are harder to quantify and less standardized than firm-level financial fundamentals. Second, the academic literature in commodities (Fama and French 2015; Gorton and Rouwenhorst 2006, among others) has long been dominated by reduced-form econometric models, largely relying on price-based inputs for one or a specific group of contracts (e.g., metals, agricultural). These models are often constrained by past returns and/or technical indicators derived from the futures curve.

And yet, commodity futures offer a highly attractive laboratory for modern predictive modeling. Unlike equities, commodity prices are tightly linked to macroeconomic conditions, exhibit abrupt regime shifts, and are influenced by distinct classes of agents, hedgers, producers, and speculators, each with different risk preferences and trading constraints. The pricing literature has already surfaced a range of theoretical factors, such as momentum, basis, hedging pressure, open interest, and skewness, which could be embedded into ML pipelines to forecast returns. Despite the diverse range of potential signals stemming from the literature on factors, most ML studies on commodities remain narrow in scope, focusing either on a single commodity group or on technical-only signals without theoretical foundations.

This chapter seeks to correct that imbalance. I follow a supervised learning framework familiar in equity ML research, where cross-sectional returns are predicted using engineered features grounded in theory and empirical evidence. I construct a wide signal database combining both theoretical premiums and traditional technical trading predictors variables. The implementation leverages a boosted tree ensemble to learn from these signals and generate daily predictions over eight target horizons across 41 commodity futures over a 30-year horizon. I then analyze the ML term structure, as shown in Blitz, Hanauer, Hoogteijling, and Howard (2023) for equity, and create a meta ensemble over the eight target horizons to mitigate target variable model risk and reduce turnover, which is usually high with machine learning models applied to financial markets.

This chapter addresses three key gaps in the financial literature. I begin by taking an expansive view of feature engineering, grounding input construction in the academic commodity factor investing literature rather than relying exclusively on off-the-shelf technical indicators. Next, I extend the equity-based machine learning methodology to the construction of long–short signals in commodity futures. Finally, I introduce the concept of ensemble portfolios by predicting over multiple target horizons, effectively characterizing the pseudo-term structure of ML signals in commodities before aggregating them into a unified strategy.

The main findings can be summarized as follows. By training gradient boosted trees in an expanding window framework, I replicate the empirical asset pricing approach seen in equities, extracting feature importance scores as pseudo-betas that quantify the relative value of each signal through time. Momentum-based indicators, particularly time-series momentum, consistently emerge as dominant features, while skewness-based signals are more prominent at shorter horizons, supporting the relevance of both trend and reversal effects. Portfolio results exhibit a mirror J-shaped pattern in annualized returns, with low correlation between ultra-short-term and long-term models. This J-shaped configuration highlights the value of horizon diversification and justifies an ensemble approach, which delivers smoother returns, lower volatility, and improved drawdown control.

This chapter is organized as follows. I begin with a review of the theoretical and empirical foundations of commodity return predictability, covering key pricing models and documented anomalies, and assess the recent literature in ML commodities. I then describe the dataset, detailing the construction of input features and the machine learning methodology used throughout the analysis. The core empirical results follow, including model-level and ensemble performance, feature relevance over time, and portfolio-level attribution. The chapter concludes with a summary of the main insights and implications for machine learning–based commodity investing.

Seeking Commodity Features: Foundations of Commodity Factor Investing

Despite their narrower universe and shorter history of financialization, commodities benefit from a surprisingly rich theoretical and empirical foundation. Early work in the economics of commodity markets introduced two core theories that remain essential today: the theory of storage (Kaldor 1939; Working 1949) and the hedging pressure hypothesis (Keynes 1930; Hirshleifer 1988). These frameworks are not academic artifacts; they provide operational foundations for return predictability in futures contracts and have shaped the classification of commodity anomalies, such as basis, term structure slope, and inventory cycles.

The theory of storage ties the slope of the futures curve to physical inventories. High inventory levels imply low convenience yields and upward-sloping curves (contango), whereas low inventories lead to backwardation, offering a reward to those willing to hold scarce assets. In contrast, the hedging pressure hypothesis focuses on the positioning of commercial hedgers and speculative traders. When commercial players dominate one side of the market, risk premiums tend to compensate the speculators taking the opposite position. These economic frictions manifest in observable features that can be engineered into factor models.

What makes commodities distinct and appealing from a quantitative perspective is their structural heterogeneity, showing even negative pairwise correlations. Energy, metals, and agricultural products are driven by different microeconomics and subject to disjoint supply shocks. And yet, empirical work has shown that a coherent set of predictors can explain return differentials across commodities. Fama and French (2015) and Gorton and Rouwenhorst (2006) established that traditional factors, such as basis and momentum, carry explanatory power in a cross-sectional framework. More importantly, these predictors are grounded in the economics of storage, risk transfer, and price discovery. The financialization of commodity markets, reflected in the growth of commodity index funds and increased institutional flows, cemented their status as an investable asset class. Part of this appeal stemmed from diversification. Commodities have historically exhibited low correlations with stocks and bonds, especially in inflationary environments. Precious metals in particular show positive co-movement with unexpected inflation (Bhardwaj, Gorton, and Rouwenhorst 2015).

Commodity Anomalies, Liquidity, and Technical Indicators

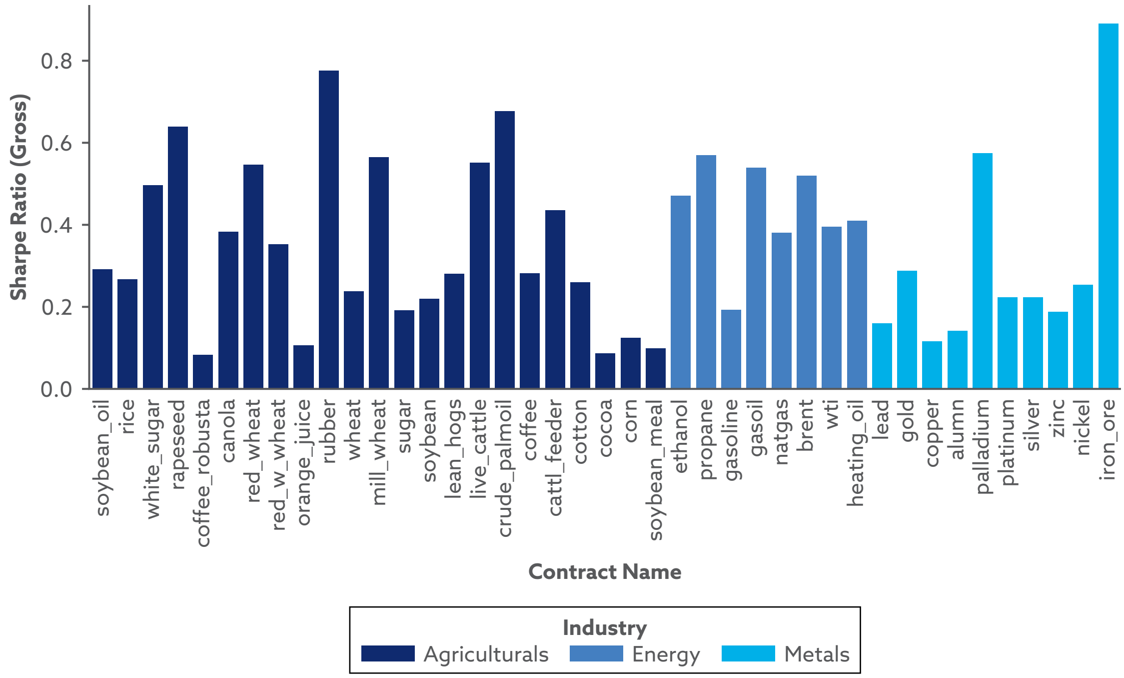
Although commodities are widely recognized for their diversification benefits, a growing body of empirical research highlights persistent return anomalies in futures markets that go beyond passive exposure. These anomalies, grounded in momentum, carry, skewness, and basis, reflect systematic inefficiencies that can be exploited through well-structured, rule-based strategies and that have been harvested as standalone quantitative investment strategies products. Building on this foundation, I constructed nine families of predictive features that span both price-based dynamics and market microstructure signals.

Momentum and time-series momentum capture recent return behavior and trend persistence. Carry and basis signals extract information from the slope and curvature of the futures curve, while basis momentum extends this insight by comparing relative performance across nearby maturities. Skewness proxies for asymmetries in return distributions that may reflect sentiment extremes or positioning risk. Additional signals include idiosyncratic volatility, which isolates contract-specific risks, and liquidity (open interest), which measures market participation and capital flows. Finally, classical technical rules, such as moving averages and relative strength index (RSI) filters, are included to benchmark modern signals against widely used technical heuristics.

Momentum: The Trend Is Your Friend

Momentum is arguably the most robust and widely validated anomaly across asset classes, and commodities are no exception. Two forms are documented: cross-sectional and time-series momentum. Cross-sectional momentum looks at how commodities perform relative to each other. The idea is simple: Go long the winners and short the losers. Miffre and Rallis (2007) showed that sorting commodities according to their past 12-month returns produces a long–short portfolio with strong and statistically significant excess returns. This effect survives standard risk adjustments, indicating that it is not only compensation for known exposures.

Exhibit 1. Time-Series Momentum for Each Futures Contract, January 1994–June 2024



Source: Bloomberg LLC.

In contrast, time-series momentum, shown in **Exhibit 1**, treats each commodity on its own. If a given contract has risen during the past year, it is likely to continue to rise, and if it has been falling, one can expect more of the same. Moskowitz, Ooi, and Pedersen (2012) implemented this logic by going long commodities with positive 12-month returns and short those with negative ones. The result was highly consistent profits across dozens of markets. This approach is the backbone of many trend-following commodity trading advisers (CTAs) and managed futures funds, proving its relevance in live strategies.

Why does momentum work? The explanations for this phenomenon range from behavioral biases, such as herding and anchoring, to structural frictions that slow information diffusion. Importantly, momentum in commodities is not a market-specific quirk. Rather, it is part of a global anomaly observed in equities, bonds, and currencies alike (Asness, Moskowitz, and Pedersen 2013).

Carry and Basis: The Term Structure as a Predictor

The term structure of futures prices offers valuable insight into the expectations and positioning of various economic agents. By examining the relationship between a commodity's futures price and its spot price or by comparing prices across different contract maturities, it is often possible to infer directional signals about future returns. This information is encapsulated in the carry or basis, which serves as a key predictive signal in many empirical asset pricing frameworks.

When futures are in backwardation (near-term futures priced below spot), it often signals supply stress and low inventories. According to the theory of storage (Kaldor 1939; Working 1949), backwardation reflects a high convenience yield, an implicit benefit to holding the physical commodity. In this context, going long earns you a premium for providing liquidity or inventory services.

Conversely, contango (when futures are above spot) reflects surplus supply and lack of urgency to hold the good. Long positions in such markets often lose value as contracts roll forward. Gorton, Hayashi, and Rouwenhorst (2013) confirmed with decades of data that commodities with tight inventories and high basis tend to earn higher returns. The hedging pressure hypothesis offers a complementary view. Commercial players, such as producers, hedge their exposure by shorting futures. Speculators willing to take the long side demand a risk premium. This dynamic leads to predictable drift in futures prices over time, typically upward, benefiting long holders. Empirical studies (e.g., de Roon, Nijman, and Veld 2000; Basu and Miffre 2013) have shown that long–short strategies based on hedging pressure or carry signals capture substantial excess returns. In practice, a simple carry strategy involves ranking commodities by their basis and going long the top quantile and short the bottom.

Skewness: When Asymmetry Matters

A more recent and nuanced anomaly is return skewness. Commodity prices are prone to spikes from geopolitical events, weather, or supply chain shocks. This tendency creates asymmetry in returns, with some markets showing fat right tails (lottery-like upside) and others, left tails (sudden crashes). Fernandez-Perez, Frijns, Fuertes, and Miffre (2018) found that this asymmetry contains predictive value. Commodities that have experienced negative skewness—frequent small gains with rare large losses—tend to deliver higher future returns. The intuition is risk based: Investors require a premium to bear downside tail risk. In contrast, positively skewed commodities (a small chance of windfall gains) attract overconfident or risk-seeking buyers, pushing prices up and expected returns down.

A skewness-based strategy, long on negatively skewed and short on positively skewed futures, earns abnormal returns not explained by momentum or carry. Although harder to implement because of higher-moment estimation and signal instability, this anomaly highlights that tail risks matter—that financial markets price the full shape of the return distribution, not only variance.

Idiosyncratic Volatility: The Role of Unexplained Risk

Idiosyncratic volatility (IVOL) refers to the volatility of a commodity's returns that is not explained by broad market or factor movements. In equity markets, high idiosyncratic volatility is often associated with lower subsequent returns (the IVOL anomaly), and a similar phenomenon has been explored in commodities. Early studies found that commodity futures with high past idiosyncratic volatility tend to underperform, suggesting a negative relation between IVOL and future returns analogous to equities' low-volatility anomaly. In the context of commodity futures, IVOL captures risks that are market specific, such as supply shocks, seasonal patterns, or logistical disruptions, that do not co-move with the broader commodity complex.

Fuertes, Miffre, and Fernandez-Perez (2015) explored the role of IVOL in commodity markets, showing that its integration into multisignal frameworks can improve portfolio diversification. For instance, combining IVOL with momentum and term structure signals allows for the construction of more robust long–short strategies by exploiting complementary sources of information. Fernandez-Perez et al. (2018) further analyzed IVOL in the presence of established commodity factors, highlighting its relevance for understanding commodity-specific risk beyond systematic drivers. Although IVOL use as a standalone signal is less straightforward, its interaction with other structural features of the market is worth investigating in an ML research context.

Basis Momentum: Riding the Curve's Slope

Basis momentum is a recently identified predictor that exploits information in the shape of the futures curve. It is defined as the difference between the momentum of near-term futures and the momentum of deferred (second-nearby) futures. This factor effectively captures the slope and curvature dynamics of the term structure: A strategy going long commodities whose nearby contracts have outperformed their second-nearby contract (and shorting those with the opposite pattern) earns significant returns. Boons and Prado (2019) introduced basis momentum and showed it strongly outperforms traditional signals, such as simple momentum or carry (basis) in predicting commodity returns. In their findings, a portfolio sorted on basis momentum produced large spread returns both in cross-section and time series, indicating that this maturity-specific momentum effect is a powerful anomaly. The basis momentum premium appears to be related to volatility and segmentation across contract maturities. For example, when speculators' risk-bearing capacity is strained, mispricing can occur between near and far contracts, which basis momentum strategies exploit. Follow-up studies have confirmed the robustness of this factor. Paschke, Prokopczuk, and Simen (2020), for instance, documented a similar "curve momentum" strategy (operating on the first two contract maturities) that achieves high Sharpe ratios and positive alpha after controlling for other factors.

Open Interest: Reading the Crowd

Open interest represents the total number of outstanding contracts (long or short) in a futures market and is often viewed as a gauge of market participation and liquidity. Uniquely, open interest is not a price-based signal but a quantity-based measure that can reflect the ease with which risk is absorbed by the market. Research by Hong and Yogo (2012) showed that movements in open interest contain valuable information about commodity risk premia. Specifically, they found that increases in open interest tend to predict lower future returns for commodity contracts, whereas declining open interest predicts higher future returns. This pattern is consistent with the idea that when more investors (particularly speculators) enter a market, driving open interest up, they provide risk-bearing capacity and push risk premiums down. Conversely, when open interest dries up, remaining hedgers must offer higher expected returns to entice speculators to take the other side. Related work also has shown that open interest is highly procyclical and correlated with macroeconomic activity, reinforcing the interpretation that it signals the broad flow of investment into commodity markets. Moreover, the predictive power of open interest holds even alongside other predictors. Cheng, Kirilenko, and Xiong (2015) found that surges in participation (often captured by open interest or investor positions) can lead to "convective risk flows," affecting pricing beyond what fundamentals alone would suggest. In practice, low-open-interest environments have been associated with subsequent positive commodity returns, as risk premiums rise to attract capital, whereas high-open-interest periods coincide with compressed returns.

Relative Strength Index: A Technical Oscillator

The relative strength index is a bounded oscillator designed to capture the velocity of recent price movements, typically using a 14-day window. In commodity futures, it is often used as a contrarian tool, with threshold levels above 70 interpreted as overbought and below 30 as oversold. Although RSI-based strategies are intuitive and widely adopted by practitioners, early academic evaluations found that the approach produced less reliable results compared with trend-following rules. Lukac, Brorsen, and Irwin (1988) reported limited profitability from standard RSI implementations across a broad set of commodity markets. More-recent studies, however, suggest that its performance can improve when combined with volume indicators or modified threshold levels. Yen and Hsu (2010) showed that hybrid strategies incorporating RSI and money flow signals perform competitively in certain commodity sectors, and Anderson and Li (2015) demonstrated that tuning RSI parameters enhances returns in energy and agricultural markets. Although the standalone signal may exhibit limited forecasting power in isolation, RSI remains a useful timing tool within multisignal frameworks, particularly for refining entries and exits in mean-reverting environments.

Moving-Average Crossovers: Trend Following in Practice

Moving-average crossover systems are among the most enduring and empirically supported technical rules in commodity futures markets. These strategies involve comparing a short-term moving average, commonly a 50-day window, with a longer-term average, such as 200 days. A buy signal is generated when the short-term average crosses above the long-term average, and a sell signal occurs on a downward crossover. This mechanism seeks to exploit persistent trends while avoiding short-term noise. Lukac et al. (1988) found that dual moving-average strategies produced statistically significant risk-adjusted returns across multiple commodity contracts. Park and Irwin (2007), in a comprehensive review of more than 90 studies, concluded that moving-average systems were consistently among the most robust technical rules in futures markets. Szakmary, Shen, and Sharma (2010) further validated the effectiveness of these rules across 28 commodity markets, showing that excess returns persist even after accounting for transaction costs and data-snooping concerns. Narayan, Ahmed, and Narayan (2015) also confirmed the signal's predictive value in a cross-sectional portfolio context, although they noted sensitivity to parameter tuning. These findings align with the broader time-series momentum literature, including the framework of Moskowitz et al. (2012), whose one-year trend filters echo the logic of long-horizon moving-average systems. Taken together, this body of work underscores the continued relevance of moving-average crossovers as reliable building blocks for systematic commodity trading strategies.

Existing Relevant Machine Learning in Commodity Markets

In this subsection, I review the most relevant recent developments in applying machine learning to commodity markets. Two key strands of literature are particularly relevant to the scope of this study.

- First, the integration of macroeconomic variables into ML-based models for commodity price forecasting has received growing attention, with numerous studies (Ben Jabeur, Khalfaoui, and Ben Arfi 2021; Wang and Zhang 2024) demonstrating the predictive value of macro-financial indicators.
- Second, a methodological distinction has emerged between univariate models, applied in isolation to individual commodities or subsets, such as metals, agricultural products, or energy, and cross-sectional (panel) approaches that model multiple commodities jointly.

I examine the main features and signals used in each stream, highlight the respective methodological trade-offs, and contrast their empirical findings. This review serves to position my own contribution, which builds long–short commodity portfolios using machine learning, in relation to these established approaches.

The Role of Macroeconomic Variables in ML-Based Commodity Forecasting Literature

Macroeconomic fundamentals are known to influence commodity markets through demand, supply, and investment channels. ML models often integrate these variables as predictors to capture broader economic signals that drive commodity price movements. Common macroeconomic features include indicators of global economic activity (e.g., industrial production growth or GDP), inflation rates, interest rates, exchange rates (especially for commodity-exporting countries' currencies), equity market indexes, and measures of liquidity or risk appetite (Gargano and Timmermann 2014; Costa, Ferreira, Gaglianone, Guillén, Issler, and Lin 2021). By including such variables, ML models aim to account for shifts in the business cycle and financial conditions that affect commodity prices. Early work by Gargano and Timmermann (2014) demonstrated the value of macro variables in commodity forecasting using traditional models. They found that certain predictors, such as commodity currency exchange rates (currencies of major commodity exporters), have significant short-horizon predictive power for commodity price indexes, while industrial production growth and investment ratios matter at longer horizons. Moreover, their results indicated that the predictive power of macroeconomic variables is regime dependent: Commodity return predictability was strongest during global recessionary periods, when macro conditions undergo substantial shift. This finding suggests that macro indicators help capture low-frequency economic trends and structural breaks.

Building on such insights, recent ML studies have incorporated large sets of macroeconomic features to improve forecast accuracy. Costa et al. (2021), for example, explored oil price forecasting with an extensive "big data" macro-financial dataset of 315 variables, combined with a suite of 23 modeling approaches, including tree-based ML (random forests, gradient boosting), regularized regressions, and classical econometric benchmarks. Their comprehensive pseudo-out-of-sample study showed that machine learning models leveraging rich macroeconomic information can significantly outperform naive benchmarks in the short run. In particular, at forecast horizons up to six months, such models as LASSO regression and tree-based ensembles (applied to macro and financial predictors alongside oil futures prices) yielded the most accurate forecasts, often achieving substantial gains in out-of-sample R^2 relative to a random walk. These improvements, on the order of two-digit percentage reductions in forecast error in some cases, underscore that ML algorithms can effectively exploit the predictive content of a broad array of macroeconomic indicators. At longer horizons (one to five years ahead), Costa et al. (2021) noted that no single ML method dominates but combinations of forecasts and structural models become more relevant, implying macro variables still contribute when used in ensemble approaches.

Other studies have confirmed that adding macroeconomic features boosts commodity price forecasts in ML frameworks. Wang and Zhang (2024) examined 22 commodity futures and showed that including both commodity-specific variables and macroeconomic factors as inputs to ML algorithms improves out-of-sample performance for the majority of those commodities. In their experiments, a gradient boosting model (LightGBM) augmented with such features as global economic indexes, interest rates, and financial market variables produced lower prediction errors than simple autoregressive benchmarks—AR(1) models—in most cases. The authors used SHAP (SHapley Additive exPlanations) values to interpret the importance of features in the ML forecasts, finding that the most influential predictors vary across commodities. For instance, some commodities are driven strongly by general macro variables (such as an aggregate demand index or currency values), whereas others are more influenced by idiosyncratic factors (such as inventory levels or past price momentum). This heterogeneity highlights that macroeconomic inputs have significant predictive value overall, but their impact can be commodity specific.

In commodity markets that are highly sensitive to policy and global conditions, macroeconomic predictors may be especially critical. For example, Ben Jabeur et al. (2021) used an explainable ML approach to predict crude oil price crashes by integrating such variables as stock market indexes, currency exchange rates, and green energy indexes. Their findings indicated that macro-financial indicators improve the early warning signals for oil market downturns by capturing broad market sentiment and structural shifts. Likewise, Ampountolas and AlGharbi (2025), using a hybrid ML model for orange juice commodity prices, reported that including financial market indexes (e.g., the S&P 500 Index) alongside macro factors improved forecast accuracy. This finding aligns with Gargano and Timmermann's (2014) emphasis on macro predictors: Broad indexes likely capture global economic conditions and investor risk appetite, which translates into better predictions of commodity demand and pricing pressures.

Univariate vs. Cross-Sectional in ML Commodity Research: Global vs. Local

A second key dimension in the literature is whether researchers model commodity prices in isolation (univariate time-series approach) or leverage information across a cross-section or panel of commodities. The univariate approach entails building a separate predictive model for each commodity's price or return, using that commodity's own lagged values and potentially some exogenous features (which could include macro variables or that commodity's specific fundamentals). In contrast, cross-sectional (or panel) approaches involve modeling multiple commodities jointly—for example, by pooling data across commodities to train a single model or by predicting the entire cross-section of commodity return at each point in time using common predictors (similar to factor investing models).

Each approach has distinct methodological trade-offs, and has been explored in recent ML-based commodity research. Univariate ML models are common in studies focusing on a particular commodity or commodity index group. Many studies take this route for major commodities, such as crude oil or gold, often comparing various ML algorithms to find the best performer for that single series. For instance, Foroutan and Lahmiri (2024) implemented 16 different machine learning and deep learning models to forecast prices of four individual markets (West Texas intermediate crude oil, Brent oil, gold, and silver), essentially treating each market as a separate forecasting task. They found that advanced neural networks, such as temporal convolutional networks or Bi-GRU networks, could outperform other models for certain commodities, while tree-based models, such as LightGBM, also performed strongly as a baseline.

The point of the univariate approach is that it allows specialized modeling of each commodity's unique dynamics, capturing specific supply shocks, seasonal patterns, or specific economic linkages (such as oil's linkage to oil inventory levels or gold's inverse relationship with interest rates) without being "diluted" by data from other commodities. Researchers can include commodity-specific features (such as metal inventories, energy rig counts, and weather variables for crops) alongside macro variables tailored to that market. This approach often improves interpretability for that commodity. One drawback, however, is that the data for each commodity are limited, which can constrain complex ML models—and most of the time, those data (like the macro variables) are lagged and their granularity is at best monthly. Overfitting is a risk, and the model might fail to generalize if the training sample (sample of one commodity) is small or if structural changes occur.

In contrast, cross-sectional and panel ML approaches attempt to "learn" across multiple commodities simultaneously, leveraging the idea that different commodities may share common patterns or factors. Cross-sectional modeling for forecasting is widespread in equities. In this approach, one might use characteristics or factors of each commodity (such as the family of factors previously introduced) at time *t* to predict the cross-section of returns in the next period. ML algorithms can be used to combine these commodity-specific characteristics in a nonlinear way.

Angelidis, Sakkas, and Tessaromatis (2025) followed this approach by comparing time-series models for individual commodities versus cross-sectional models that predict all commodities' returns jointly. They examined 37 commodity futures and found that cross-sectional models produce superior forecasts compared with time-series models for commodity returns. In other words, a model that pools information across commodities (and exploits cross-sectional predictors, such as each commodity's basis, momentum, or other characteristics at a given time) achieved lower overall prediction error than modeling each commodity in isolation. The authors reported that combining forecasts and using modern ML methods further enhanced performance, suggesting that the cross-sectional signals contain exploitable structure that ML methods can capture more effectively relative to traditional approaches. Intuitively, cross-sectional models benefit from wider information sets. They can detect, for example, that commodities with certain traits (say, a high prior-year return or a negative skewness) tend to mean-revert or continue trending, based on patterns learned from the full panel of commodities. This dynamic can improve generalization because the model draws on a larger sample of observations (multiple commodities over time) to learn predictor–response relationships. There are several methodological trade-offs between univariate and cross-sectional approaches. Generalization versus specialization is a key consideration.

Cross-sectional ML models effectively assume that different commodities' price dynamics share some underlying functional form or factors, which the model can learn. This pooling can dramatically increase the effective sample size for training, aiding generalization and reducing estimation noise. Indeed, the superiority of the cross-sectional ML in Angelidis et al. (2025) echoes findings in other asset classes (e.g., equities) that "global" models using panel data can outperform many separate "local" models by borrowing strength across series. Nonetheless, the optimal approach may depend on the context: If the forecasting goal is to rank or allocate across commodities (long–short portfolio construction), panel ML models are naturally suited. If the goal is to predict exact price movements of a particular commodity (for hedging purposes), however, a univariate model with that commodity's key drivers might be better suited.

Empirical Results

This section presents the empirical evaluation of the proposed framework. I begin by describing the dataset, feature construction, and methodological setup used for forecasting across multiple horizons. Next, I analyze the information content and relative importance of each feature within boosted tree–based models to understand how predictive drivers vary with horizon length. Finally, I assess the out-of-sample performance of individual models and ensemble approaches through long–short portfolio simulations, comparing their risk-adjusted returns and analytics across forecast horizons.

Data and Methodology

The sample universe consists of 41 daily continuous futures contracts: 23 agricultural commodities, 8 energy products, and 10 metals, as shown in **Exhibit 2**. I sourced prices and open interest data from Bloomberg for the first and second contracts along the term structure, which were used to construct the features. For each contract, I computed a continuous price series using the back-adjusted ratio methodology. This approach applies a multiplicative adjustment five days prior to each contract roll, using the ratio between the next and previous contract prices to rescale all preceding prices. This process ensures that returns and all price-based features remain consistent across the series, avoiding artificial jumps caused by roll effects. For return-based features, all prices not denominated in US dollars were converted to US dollars. When I computed price ratios, however, contracts were maintained in their local currency to minimize the influence of exchange rate movements on daily price dynamics.

As outlined in the previous section, I constructed the feature set on the basis of both theoretical and practical insights from the literature. Theoretically motivated features include carry, time-series momentum, cross-sectional momentum, basis momentum, and idiosyncratic volatility, which are well documented in empirical asset pricing studies. Practically motivated signals were drawn from widely used technical analysis heuristics, such as moving average crossovers, the relative strength index, and moving averages of open interest, reflecting strategies often implemented by systematic traders. This process resulted in a total of nine distinct feature families, each aiming to capture different structural aspects of commodity price behavior.

In line with standard practices in equity-based machine learning research and inspired by the lookback window parameter (J) in Jegadeesh and Titman (2001), I applied 14 different lookback periods to each of the nine feature types. The lookbacks, expressed in trading days, are 10, 22, 44, 66, 88, 130, 200, 252, 300, 382, 400, 504, 600, and 756. These windows are designed to span ultra-short-term, short-term, medium-term, long-term, and ultra-long-term horizons, allowing the models to learn from trend dynamics and seasonal patterns. This approach yielded a total of 126 distinct daily features per contract.

The full dataset consists of approximately 40 million data points across contracts, dates, and features. For supervised machine learning, I computed daily returns from continuous futures

Exhibit 2. Investment Universe

Group Level 1	Group Level 2	Contract/Exchange Name	Group Level 1	Group Level 2	Contract/Exchange Name
Agriculturals	grains_oilseeds	Canola (ICE Canada)	Energy_fut	gas	Natural Gas (NYMEX – CME Group)
Agriculturals	grains_oilseeds	Corn (CBOT – CME Group)	Energy_fut	gas	Propane – Mt Belvieu LDH (NYMEX – CME Group)
Agriculturals	grains_oilseeds	Crude Palm Oil (Bursa Malaysia Derivatives – BMD)	Energy_fut	oil	Brent Crude Oil (ICE Europe)
Agriculturals	grains_oilseeds	Milling Wheat (Euronext – MATIF)	Energy_fut	oil	Ethanol (CBOT – CME Group)
Agriculturals	grains_oilseeds	Rapeseed (Euronext)	Energy_fut	oil	Low Sulfur Gasoil (ICE Europe)
Agriculturals	grains_oilseeds	Rough Rice (CBOT – CME Group)	Energy_fut	oil	NY Harbor ULSD – Ultra Low Sulfur Diesel (NYMEX – CME Group)
Agriculturals	grains_oilseeds	Soybean Meal (CBOT – CME Group)	Energy_fut	oil	RBOB Gasoline (NYMEX – CME Group)
Agriculturals	grains_oilseeds	Soybean Oil (CBOT – CME Group)	Energy_fut	oil	WTI Crude Oil (NYMEX – CME Group)
Agriculturals	grains_oilseeds	Soybeans (CBOT – CME Group)	Metals	base_metals	Copper (COMEX – CME Group)
Agriculturals	grains_oilseeds	Wheat – Hard Red Winter (KCBT – CME Group)	Metals	base_metals	Iron Ore 62% Fe CFR China (SGX)
Agriculturals	grains_oilseeds	Wheat – Red Spring (MGEX – CME Group)	Metals	base_metals	Lead (LME)
Agriculturals	grains_oilseeds	Wheat – Soft Red Winter (CBOT – CME Group)	Metals	base_metals	Nickel (LME)

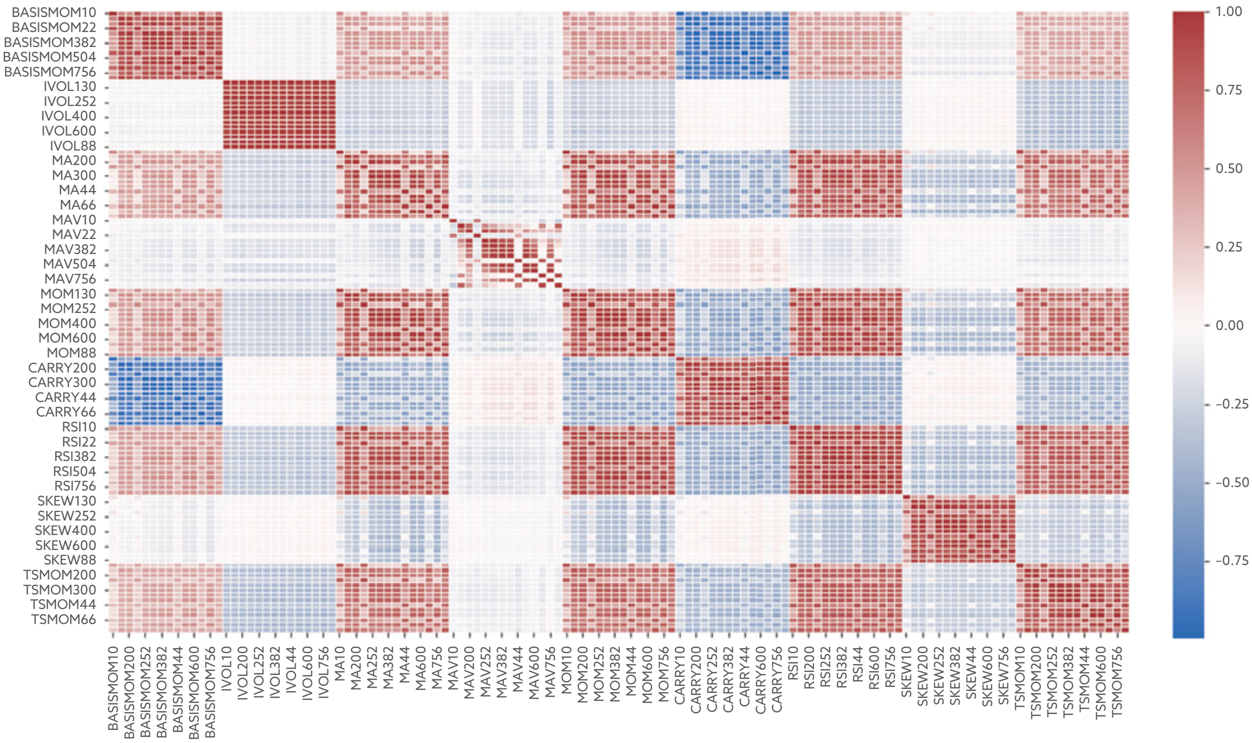
Group Level 1	Group Level 2	Contract/Exchange Name	Group Level 1	Group Level 2	Contract/Exchange Name
Agriculturals	livestock	Feeder Cattle (CME)	Metals	base_metals	Primary Aluminum (LME)
Agriculturals	livestock	Lean Hogs (CME)	Metals	base_metals	Zinc (LME)
Agriculturals	livestock	Live Cattle (CME)	Metals	precious_metals	Gold 100 oz (COMEX – CME Group)
Agriculturals	softs	Cocoa (ICE US)	Metals	precious_metals	Palladium (NYMEX – CME Group)
Agriculturals	softs	Coffee Arabica (ICE US)	Metals	precious_metals	Platinum (NYMEX – CME Group)
Agriculturals	softs	Coffee Robusta (ICE Europe)	Metals	precious_metals	Silver (COMEX – CME Group)
Agriculturals	softs	Cotton No.2 (ICE US)			
Agriculturals	softs	Frozen Concentrated Orange Juice (ICE US)			
Agriculturals	softs	Rubber (OSE – Osaka Exchange)			
Agriculturals	softs	Sugar #11 – Raw Sugar (ICE US)			
Agriculturals	softs	Sugar #5 – White Sugar (ICE Europe)			

Source: Bloomberg LLC.

price series over various prediction horizons: 5, 10, 22, 44, 66, 80, 100, and 200 days. Each prediction horizon corresponds to a separate supervised learning model, enabling the analysis of time-varying predictability across horizons. All features and target variables are normalized into percentiles to ensure comparability across commodities and over time, mitigating scale differences and enhancing model stability during training. **Exhibit 3** shows the correlation heatmap of percentile-transformed features. It reveals that several feature families, such as BASISMOM, CARRY, and SKEW, are largely uncorrelated with each other, indicating they capture distinct dimensions of return predictability. Within individual families, the impact of lookback parameters varies: IVOL features are highly correlated across different horizons, suggesting that the choice of lookback has limited influence on their information content, while such features as MOM, TSMOM, and RSI show greater variation with horizon length, highlighting their sensitivity to the lookback parameter.

Because the scope of this chapter is not to compare different supervised models but rather to focus on the application of ML to commodities, I selected a single type of ML model—a regression boosted tree model—to train over all the various prediction horizons. Boosted trees are a powerful and regularized implementation of gradient boosting, widely used in cross-sectional finance (see Guida and Coqueret 2019). Instead of fitting one large tree, the algorithm builds an ensemble of shallow trees, where each new tree focuses on correcting the residuals of the

Exhibit 3. Correlation Heatmap of Percentile-Transformed Features over the Full Sample



Source: Bloomberg LLC.

previous ones. This sequential learning process enables the model to capture complex, nonlinear interactions between predictors and the target variable. At each iteration, boosted trees minimize a regularized objective function that combines a differentiable loss, such as squared error or logistic loss, with penalties on model complexity, including the number of leaves and the magnitude of leaf weights. The algorithm leverages both the gradient and the Hessian of the loss function to construct an accurate approximation of the optimal tree structure. Split quality is evaluated using a second-order Taylor expansion, which guides the model toward partitions that effectively reduce the loss while maintaining parsimony. These technical refinements make boosted trees especially effective in financial settings, where predictive signals are often subtle and embedded in noisy data.

Each boosted tree model is trained using an expanding window approach, beginning with a minimum of three years of historical data. For the initial training window of five years, 90% of the observations are used for training and the remaining 10%, corresponding to approximately six months, are reserved for out-of-sample testing. In subsequent iterations, the test set is kept fixed to ensure that the training data do not become too temporally distant from the prediction period. Models are retrained annually by extending the training window to incorporate the most recent data, while maintaining a fixed-length, non-overlapping six-month test set. To avoid look-ahead bias, particularly in models that predict long-term returns (such as the 200-day forward horizon), a temporal buffer is introduced. Specifically, the end date of the test features is shifted to ensure that the return label window does not extend into the subsequent prediction period. For example, if the next prediction period begins in January 2023, the test features for a 200-day horizon model would end in January 2022. This adjustment isolates the training and testing phases and prevents any leakage of future information.

Because some hyperparameters introduce randomness, each model is run 20 times with different random seeds for each target variable. The results are then averaged to reduce statistical noise and improve robustness. Consistent with standard practice in financial machine learning, I treat the model's output not as a point forecast but as a ranking signal. Each supervised model produces daily predictions across assets, which I interpret as cross-sectional scores. These scores are then normalized into quantiles and used to sort assets into a long–short portfolio, in a manner closely aligned with traditional factor-based investing. The same score drives both portfolio construction and weighting: Assets with the highest predicted returns are assigned to the long leg, and those with the lowest scores populate the short leg. Within each leg, position sizes are proportional to the scores themselves, giving more weight to the most confident predictions. This setup ensures that the model's relative convictions are fully reflected in the portfolio. I fix the gross exposure of each leg at 1, resulting in a total gross exposure of 2. The process is repeated independently for each prediction horizon, enabling horizon-specific portfolio views.

Model hyperparameters are tuned separately for each target horizon and rebalancing point via a light grid search (for each hyperparameter, I selected three variants), choosing the best parameter set that minimizes the mean squared error. The grid search explores tree depth (from 1 to 3), learning rate (from 0.05 to 0.005), number of trees (from 50 to 150), feature subsampling per tree (from 0.55 to 0.85), and feature subsampling per level (from 0.55 to 0.85), balancing model complexity with out-of-sample robustness. The L2 regularization parameter is held at its default value of 1.

Results: Interpreting Feature Importance across Horizons and Models

In linear models, coefficients represent the marginal contribution of each predictor to the target variable. Although tree-based models, such as gradient boosting, do not yield coefficients in the classical sense, they do provide feature importance scores that reflect the relative contribution of each feature to the model's predictions. These scores can be derived using various metrics, such as the following:

- *Gain:* the improvement in the model's loss function brought by a feature when it is used to split the data
- *Cover:* the number of samples affected by splits on that feature
- *Frequency:* how often a feature is used in splits across all trees

Alternatively, SHAP (SHapley Additive exPlanations) values (Shapley 1953) offer a game-theoretic approach to feature attribution. SHAP values assign each feature a contribution value for individual predictions, enabling both local (per-sample) and global (aggregated) interpretability. Unlike gain or frequency, SHAP values are consistent and provide additive explanations across features. As Simonian, Wu, Itano, and Narayanam (2019) proposed, feature importance can be interpreted as pseudo-betas, offering a directionless yet informative measure of marginal predictive power. This concept plays a central role in interpreting machine learning models, often (wrongly) seen as opaque, and offers insights into which signals the model relies on and how this reliance evolves through time. Understanding which characteristics influence model predictions directly relates to the interpretability challenges discussed in empirical asset pricing applications using machine learning (Gu, Kelly, and Xiu 2020).

In **Exhibit 4**, I report the top 10 normalized feature importance values for each model based on different target horizons, using a last-10-year window for clarity's sake. Each table is sorted by the cumulative importance of each feature across the last 10 years. On average, the top 10 features account for 20% of total importance, while the top 50 features explain roughly 66%. Because of regularization constraints, on average about 40 of the 126 input features are consistently unused by the models.

The models are trained in an expanding window framework, meaning that feature importance in a given year reflects cumulative learning up to that point. For instance, importance values observed for end-of-year 2020 models include data from January 1994 to May 2020, yet meaningful shifts occur in model behavior following the onset of the COVID-19 crisis. This effect is especially pronounced in short-horizon models, such as the 5-day and 10-day variants. As more postcrisis data enter the training set, these models adjust rapidly. In contrast, longer-horizon models, such as the 200-day version, exhibit greater stability, with little change observed between 2019 and 2020. This behavior reflects the model's regularization bias toward persistent signals and its aversion to unstable feature dynamics.

One of the most consistent findings across models and timeframes is the dominance of momentum-based signals, particularly time-series momentum. TSMOM252 appears in the top 10 features for every model, and TSMOM300 is ranked similarly in six out of eight models. Although validating the factor investing literature is not the primary objective of this chapter, the results, derived from a purely data-driven boosted tree model, confirm the predictive power

Exhibit 4. Yearly Normalized Features' Importance

Feature	Target_Variable	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
MOM252	*5d*	2.3%	2.3%	5.8%	4.8%	2.5%	9.3%	2.3%	6.0%	3.4%	7.0%	2.3%
SKEW130	*5d*	1.5%	1.5%	2.3%	2.1%	1.6%	4.4%	1.7%	6.2%	3.8%	7.9%	2.4%
TSMOM252	*5d*	1.9%	1.9%	3.8%	3.1%	1.8%	4.9%	1.9%	4.1%	2.5%	3.7%	1.6%
BASISMOM130	*5d*	1.5%	1.6%	2.5%	3.4%	1.8%	4.0%	2.0%	3.9%	2.4%	5.0%	1.7%
MOM88	*5d*	2.4%	2.0%	5.2%	1.8%	2.0%	4.5%	1.9%	3.7%	1.9%	2.4%	1.7%
SKEW200	*5d*	1.5%	1.5%	2.3%	2.1%	1.4%	3.9%	1.6%	4.4%	2.1%	5.3%	2.0%
TSMOM300	*5d*	1.6%	1.4%	4.2%	2.2%	1.5%	4.1%	1.6%	2.9%	1.6%	3.3%	1.4%
SKEW400	*5d*	1.4%	1.4%	2.3%	1.5%	1.2%	1.9%	1.1%	3.3%	2.4%	4.6%	1.5%
CARRY66	*5d*	1.3%	1.4%	2.4%	2.2%	1.5%	2.5%	1.3%	2.1%	2.0%	3.3%	1.3%
MOM756	*5d*	1.2%	1.3%	1.9%	1.9%	1.3%	2.9%	1.3%	2.5%	1.6%	3.0%	1.3%

Feature	Target_Variable	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
MOM252	*10d*	2.4%	2.4%	6.7%	5.4%	3.1%	3.3%	2.2%	4.0%	2.7%	5.5%	1.9%
TSMOM252	*10d*	1.9%	1.9%	3.8%	3.8%	3.4%	3.2%	2.3%	4.8%	2.8%	3.3%	1.7%
SKEW130	*10d*	1.2%	1.2%	2.5%	3.1%	1.8%	2.0%	1.6%	5.1%	2.9%	6.4%	2.1%
BASISMOM130	*10d*	1.6%	1.5%	3.6%	4.2%	2.6%	2.3%	2.0%	3.4%	2.4%	3.9%	1.5%
SKEW200	*10d*	1.3%	1.3%	2.0%	2.2%	1.8%	2.0%	1.6%	3.5%	2.7%	6.0%	1.9%
TSMOM300	*10d*	2.1%	2.1%	4.5%	2.9%	2.0%	2.4%	1.7%	2.3%	1.8%	2.4%	1.6%
SKEW382	*10d*	1.4%	1.4%	2.6%	1.9%	1.4%	1.3%	1.1%	2.7%	3.3%	6.6%	1.6%
BASISMOM66	*10d*	1.4%	1.4%	3.1%	2.9%	1.8%	1.8%	1.2%	2.0%	1.8%	3.8%	1.2%
SKEW756	*10d*	1.3%	1.5%	3.1%	2.5%	2.2%	2.1%	1.7%	2.6%	2.3%	0.0%	1.5%
SKEW600	*10d*	1.4%	1.0%	3.1%	2.4%	1.8%	1.5%	1.3%	2.5%	2.5%	2.1%	1.3%

Feature	Target_Variable	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
MOM252	*22d*	2.7%	2.7%	5.5%	5.6%	2.2%	2.9%	2.1%	3.5%	5.4%	6.3%	2.0%
TSMOM252	*22d*	2.4%	2.5%	6.5%	5.5%	2.4%	3.6%	2.5%	4.6%	4.2%	3.4%	2.2%
SKEW130	*22d*	1.6%	1.7%	3.4%	2.9%	1.7%	2.0%	2.0%	5.6%	7.5%	8.8%	2.3%
SKEW200	*22d*	1.6%	1.6%	1.7%	2.4%	1.5%	2.0%	1.8%	4.0%	5.8%	8.2%	2.2%
TSMOM300	*22d*	2.3%	2.8%	4.7%	3.9%	1.6%	2.9%	2.2%	2.0%	3.2%	3.9%	1.4%
SKEW252	*22d*	0.0%	0.9%	0.0%	0.0%	1.6%	1.9%	1.4%	4.7%	7.9%	8.3%	2.0%
BASISMOM130	*22d*	1.3%	1.5%	3.0%	3.1%	1.8%	2.0%	1.9%	2.7%	3.9%	3.6%	1.6%
MOM756	*22d*	1.3%	1.4%	2.7%	2.4%	1.5%	1.8%	1.7%	2.7%	3.9%	4.3%	1.8%
SKEW382	*22d*	1.8%	1.8%	2.2%	2.2%	1.5%	1.9%	1.7%	3.3%	5.1%	0.0%	1.2%
BASISMOM200	*22d*	1.4%	1.3%	0.0%	0.0%	1.4%	1.5%	1.5%	3.4%	4.5%	4.2%	1.4%

Feature	Target_Variable	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
TSMOM252	*44d*	2.6%	2.8%	5.6%	3.5%	3.8%	2.8%	2.9%	3.1%	2.2%	3.6%	2.0%
MOM252	*44d*	2.3%	2.6%	5.8%	3.4%	3.1%	2.5%	1.5%	2.0%	1.8%	5.3%	1.9%
BASISMOM200	*44d*	1.7%	1.5%	3.8%	2.8%	3.0%	2.4%	2.2%	2.9%	2.2%	5.8%	2.0%
SKEW252	*44d*	1.8%	0.9%	0.0%	2.6%	2.8%	1.7%	1.9%	3.4%	3.1%	9.4%	2.7%
BASISMOM130	*44d*	1.4%	1.3%	2.5%	2.2%	2.7%	2.2%	2.4%	2.5%	2.1%	5.5%	1.9%
TSMOM300	*44d*	2.0%	2.0%	5.0%	2.5%	2.5%	1.9%	2.3%	1.9%	1.6%	3.7%	1.3%
BASISMOM88	*44d*	1.6%	1.5%	3.5%	2.4%	3.0%	2.0%	2.3%	2.5%	1.8%	3.3%	1.9%
SKEW130	*44d*	1.3%	1.6%	2.5%	0.0%	2.2%	1.8%	2.0%	2.9%	2.3%	6.1%	2.5%
CARRY88	*44d*	1.7%	1.6%	2.8%	2.4%	3.0%	2.0%	2.2%	2.0%	1.1%	3.2%	1.9%
SKEW756	*44d*	1.6%	1.5%	3.8%	2.0%	2.3%	1.6%	1.7%	2.0%	2.1%	3.2%	1.9%

Feature	Target_Variable	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
TSMOM252	*66d*	2.6%	3.2%	5.5%	3.3%	3.3%	3.0%	2.7%	2.8%	2.3%	4.3%	2.0%
SKEW252	*66d*	1.4%	1.3%	0.0%	1.5%	2.2%	1.4%	1.8%	2.9%	3.8%	10.5%	3.1%
BASISMOM130	*66d*	1.1%	1.7%	2.8%	2.0%	3.0%	2.2%	2.3%	2.6%	2.5%	6.8%	2.2%
BASISMOM200	*66d*	1.6%	1.8%	3.6%	2.5%	2.7%	2.5%	2.1%	2.3%	2.0%	6.4%	1.8%
SKEW300	*66d*	1.3%	1.7%	2.4%	1.5%	1.8%	2.1%	2.0%	2.4%	2.4%	7.3%	1.9%
SKEW130	*66d*	1.2%	1.3%	3.2%	2.0%	1.8%	2.0%	1.9%	2.7%	2.7%	5.3%	2.2%
MOM252	*66d*	2.0%	2.9%	5.0%	2.9%	0.0%	2.0%	1.2%	1.6%	1.6%	4.6%	1.6%
TSMOM300	*66d*	1.9%	2.4%	4.1%	1.6%	1.7%	2.5%	3.0%	1.4%	1.6%	3.2%	1.6%
SKEW600	*66d*	1.6%	1.9%	3.7%	2.1%	2.4%	1.6%	1.9%	2.3%	2.1%	3.5%	1.6%
SKEW400	*66d*	2.1%	2.3%	3.9%	2.0%	1.8%	1.7%	1.8%	2.2%	2.4%	3.1%	1.3%

Feature	Target_Variable	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
TSMOM252	*88d*	3.1%	5.8%	6.5%	3.4%	4.1%	2.9%	2.8%	2.4%	1.8%	3.3%	1.6%
TSMOM200	*88d*	3.1%	5.2%	6.0%	3.4%	4.1%	2.3%	2.3%	2.3%	1.9%	3.8%	1.6%
SKEW130	*88d*	1.7%	2.3%	3.7%	1.9%	2.5%	2.0%	2.1%	3.3%	3.2%	7.9%	2.4%
BASISMOM130	*88d*	1.8%	2.3%	3.2%	2.1%	2.7%	2.0%	2.1%	3.2%	2.9%	5.8%	2.1%
BASISMOM200	*88d*	1.7%	2.5%	3.5%	2.7%	3.3%	2.1%	2.5%	2.6%	1.9%	5.3%	1.7%
SKEW200	*88d*	1.7%	2.2%	2.5%	2.2%	2.7%	2.0%	2.0%	2.1%	1.8%	8.2%	1.8%
SKEW252	*88d*	1.2%	0.0%	0.0%	0.0%	1.8%	1.6%	2.4%	3.1%	3.5%	9.1%	2.6%
IVOL400	*88d*	1.7%	2.1%	2.6%	1.9%	2.2%	1.9%	1.9%	2.0%	2.1%	4.3%	1.8%
TSMOM400	*88d*	2.2%	4.0%	5.7%	3.3%	3.3%	1.1%	0.0%	1.7%	1.5%	0.0%	1.0%
SKEW600	*88d*	2.5%	2.9%	3.9%	1.9%	2.3%	1.4%	1.7%	2.2%	2.3%	0.0%	1.4%

Exhibit 4. Yearly Normalized Features' Importance *(continued)*

Feature	Target_Variable	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
TSMOM252	*100d*	3.3%	4.5%	6.1%	3.4%	7.0%	2.5%	2.3%	1.9%	1.7%	3.2%	1.5%
TSMOM200	*100d*	3.2%	4.3%	5.6%	3.3%	5.9%	2.4%	2.2%	2.1%	1.8%	3.4%	1.8%
SKEW130	*100d*	2.0%	2.0%	3.2%	2.2%	3.7%	2.1%	2.2%	3.4%	3.2%	8.4%	2.7%
BASISMOM130	*100d*	2.1%	1.9%	3.3%	2.2%	4.6%	2.1%	2.4%	3.3%	2.6%	5.6%	2.3%
BASISMOM200	*100d*	1.9%	1.9%	3.1%	2.9%	5.2%	2.3%	2.4%	2.4%	1.9%	3.3%	1.9%
SKEW200	*100d*	1.7%	1.7%	2.2%	2.2%	3.7%	2.0%	2.0%	1.9%	1.8%	7.6%	2.0%
TSMOM400	*100d*	2.6%	2.8%	5.5%	3.6%	5.3%	1.0%	0.0%	1.2%	1.5%	2.4%	1.6%
IVOL400	*100d*	1.8%	2.0%	2.6%	1.9%	3.6%	1.7%	1.7%	1.9%	2.1%	3.9%	1.9%
SKEW600	*100d*	2.4%	2.9%	3.8%	2.1%	3.2%	1.6%	1.8%	2.2%	2.1%	0.0%	1.5%
SKEW400	*100d*	2.9%	2.4%	4.0%	1.9%	3.4%	1.8%	1.8%	2.2%	1.6%	0.0%	1.5%

Feature	Target_Variable	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
TSMOM200	*200d*	3.2%	6.2%	6.1%	6.9%	6.5%	3.0%	2.9%	2.6%	2.4%	6.3%	2.3%
TSMOM252	*200d*	2.5%	7.9%	6.6%	7.2%	5.4%	2.4%	2.6%	1.8%	1.6%	5.3%	1.9%
IVOL300	*200d*	1.6%	3.5%	3.7%	4.8%	3.6%	2.0%	2.1%	1.4%	1.2%	5.3%	1.7%
IVOL382	*200d*	1.6%	2.8%	3.6%	4.1%	3.9%	2.0%	2.2%	2.0%	1.6%	3.7%	1.5%
TSMOM300	*200d*	3.1%	7.0%	5.6%	5.6%	1.9%	1.3%	0.0%	1.4%	1.4%	0.0%	1.5%
SKEW600	*200d*	1.9%	3.3%	3.2%	3.3%	2.9%	1.6%	1.3%	1.4%	1.1%	4.1%	1.8%
IVOL252	*200d*	1.9%	3.9%	3.8%	3.9%	3.6%	1.7%	1.6%	1.3%	1.4%	0.0%	1.6%
MOM600	*200d*	1.5%	0.0%	3.0%	2.8%	2.7%	2.1%	1.9%	1.8%	1.8%	4.4%	2.2%
MA756	*200d*	1.6%	2.1%	2.1%	2.3%	2.2%	1.8%	1.5%	1.4%	1.8%	4.7%	2.2%
TSMOM382	*200d*	2.1%	4.5%	5.6%	5.6%	0.0%	1.2%	1.1%	1.1%	1.1%	0.0%	1.2%

Feature	Target_Variable	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
TSMOM252	*ens*	2.3%	3.4%	4.9%	3.7%	3.5%	2.8%	2.2%	2.8%	2.1%	3.3%	1.6%
SKEW130	*ens*	1.3%	1.3%	2.3%	1.6%	1.7%	2.0%	1.7%	3.4%	3.0%	6.1%	2.0%
BASISMOM130	*ens*	1.3%	1.6%	2.6%	2.5%	2.5%	2.1%	1.9%	2.6%	2.2%	4.0%	1.6%
MOM252	*ens*	1.7%	1.4%	3.5%	3.0%	1.2%	2.7%	1.2%	1.9%	1.7%	3.5%	1.3%
SKEW200	*ens*	1.5%	1.6%	1.8%	1.7%	1.8%	1.9%	1.5%	2.4%	2.2%	4.4%	1.6%
BASISMOM200	*ens*	1.3%	1.3%	2.1%	1.6%	2.4%	1.6%	1.6%	2.3%	2.0%	4.0%	1.5%
TSMOM300	*ens*	2.0%	2.3%	3.1%	2.3%	1.4%	2.0%	1.6%	1.6%	1.5%	2.5%	1.3%
SKEW600	*ens*	1.6%	1.8%	2.9%	1.9%	1.8%	1.6%	1.3%	1.7%	1.8%	1.3%	1.4%
TSMOM200	*ens*	1.3%	2.1%	2.4%	1.9%	2.4%	1.6%	1.5%	1.2%	1.1%	2.3%	1.2%
SKEW252	*ens*	0.7%	0.5%	0.0%	0.6%	1.3%	1.2%	1.5%	2.7%	3.1%	5.5%	1.9%

Source: Bloomberg LLC.

of trend signals in commodity markets. Cross-sectional momentum (MOM252) also appears frequently among the most important features, suggesting that both absolute and relative price dynamics are integral to model performance.

Another notable result is the prominence of skewness-based features, particularly in short-term models. For example, SKEW130 and SKEW200 display elevated importance in 2022, especially in models with shorter forecast horizons. This finding is consistent with the hypothesis that negative skewness can trigger short-term price overreactions, followed by reversals (Fernandez-Perez et al. 2018). In this sense, short-term models tend to load more on reversal-type signals, often combining skewness with momentum indicators, whereas longer-horizon models gravitate more heavily toward pure trend-following signals, such as TSMOM.

Finally, it is worth noting that classical technical indicators, such as RSI, MA, and MAV, rank lower in importance for almost all horizons. Regardless of lookback period, these features appear less frequently in the top 10, suggesting that when competing against more academically grounded signals, their marginal contribution to predictive accuracy is limited.

Performance Analysis Comparing Models and Ensemble

Exhibit 5 presents the detailed statistics of long–short portfolios constructed for each target horizon. Each leg of the long–short portfolio is weighted according to its normalized prediction score. Given the narrow nature of the commodity investment universe, limited to only

Exhibit 5. Performance and Analytics Table for Models and Ensemble

Model	Annualized Return Net of Transaction Costs	Annualized Volatility	Sharpe	Sortino	Downside Volatility	Value at Risk (95%)	Success Rate	Max. Drawdown	Daily Turnover
5d	49%	13.5%	3.6	5.3	9.2%	1.2%	60%	−17%	38%
10d	39%	13.5%	2.9	4.2	9.1%	1.2%	58%	−24%	37%
22d	26%	13.5%	1.9	2.7	9.7%	1.3%	56%	−28%	31%
44d	20%	13.8%	1.5	2.1	9.8%	1.3%	55%	−26%	29%
66d	18%	13.9%	1.3	1.8	9.9%	1.3%	55%	−34%	27%
88d	18%	13.7%	1.3	1.9	9.6%	1.3%	54%	−25%	28%
100d	20%	13.7%	1.4	2.0	9.8%	1.3%	55%	−25%	27%
200d	22%	13.1%	1.7	2.4	9.1%	1.2%	55%	−23%	27%
Ensemble	27%	10.9%	2.4	3.3	8.3%	1.0%	57%	−22%	17%

Source: Bloomberg LLC.

41 contracts, it would not be meaningful to form decile portfolios. Instead, the long portfolio includes contracts with normalized predictions above the cross-sectional median, and the short portfolio contains those below it. As explained previously, all models are retrained each year in December and new predictions are made every day for the following year that becomes the out-of-sample results. Hence, the models' portfolio would change daily according to new unseen features' input. I applied a generic prediction-level weighting scheme (comparable to a factor intensity) to accentuate the role of predictions in the performance of the models' long–short portfolio.

Results are shown for the full sample of out-of-sample predictions from January 2000 until June 2024, net of transaction costs, which are computed, for the sake of simplicity, as 5 bps per unit of turnover. The ensemble is created using the average of normalized predictions, which are normalized once more to filter and weight the assets in the same fashion as the other target horizon models. The gross market value (GMV) is kept at a constant level for each model because a new model weight is recomputed each day with new predictions between two yearly retrainings.

Annualized net performance by target horizon exhibits a nonmonotonic pattern: Performance is higher for shorter-term horizons—it decreases from 5 to 66 days—and it rises slightly after 88 days. Sharpe and Sortino ratios are generally high, especially for shorter-term models (Sharpe ratio of 3.6 and 2.9 for, respectively, 5d and 10d). The cost, however, is a high level of turnover. Even with conservative transaction costs of 5 bps, the capacity (maximum assets under management that is tradeable without impacting the market too much and leaving too visible a market footprint) is in the low hundreds of millions, especially for achieving the required turnover on such contracts as orange juice, rubber, or rice. The success rate, defined as the number of positive returns over the total number of daily returns, is generally very good for all models and very high for shorter-term horizon strategies (60.0% for 5d, 58.3% for 10d).

The risk measures do not indicate large differences among the models, with volatility being around 13% and downside volatility being below 10%. There are some differences in maximum drawdown over the full period, part of which occurred during the COVID-19 pandemic; maximum drawdown is high in general and in particular for the 22d and 66d models. For these models, maximum drawdown exceeds the rule of thumb of 1.5 times the volatility of the strategy, which is commonly used as a "red flag" in multistrategy hedge funds pods. Regarding the ensemble, the results show that blending predictions together provides a risk diversification effect that smooths volatility and tail risk. While providing a mid-range net annualized return of 27.3%, the ensemble cuts annual volatility by 2 to 3 points relative to the best standalone models, lowers value at risk, and keeps the drawdowns moderate.

Exhibit 6 shows the full-period pairwise correlation matrix for the strategies, each defined by a specific target horizon, ranging from short term (5d) to long term (200d), including their relationship to an ensemble strategy that aggregates them.

First, the matrix reveals a high degree of correlation among the medium-term models (22d to 100d), with values typically exceeding 0.80. For example, the correlation between the 22d and 44d strategies is 0.84, while the 66d and 88d correlation reaches 0.88. These elevated correlations suggest that the corresponding models are identifying overlapping patterns in commodity returns, arising from important common features. Although this finding confirms the robustness of certain signals, it also implies a degree of informational redundancy among these models, which may limit the incremental benefit when combining them into the ensemble.

Conversely, the 200d strategy stands out because of its markedly lower correlations. For instance, its correlations with the 5d and 22d strategies are only 0.41 and 0.49, respectively. This divergence suggests that the 200d model is extracting fundamentally different signals, as depicted by the top 10 most important features table (Exhibit 4), potentially aligned with slow-moving structural or seasonal phenomena. In an ensemble framework, such low

Exhibit 6. Correlation Matrix for Long–Short Portfolios

Target Variable	5d	10d	22d	44d	66d	88d	100d	200d	Ensemble
5d	1.00	0.83	0.77	0.73	0.71	0.66	0.65	0.41	0.83
10d	0.83	1.00	0.83	0.78	0.76	0.71	0.69	0.50	0.88
22d	0.77	0.83	1.00	0.84	0.81	0.74	0.72	0.49	0.89
44d	0.73	0.78	0.84	1.00	0.86	0.79	0.77	0.51	0.91
66d	0.71	0.76	0.81	0.86	1.00	0.88	0.84	0.58	0.93
88d	0.66	0.71	0.74	0.79	0.88	1.00	0.92	0.63	0.91
100d	0.65	0.69	0.72	0.77	0.84	0.92	1.00	0.64	0.90
200d	0.41	0.50	0.49	0.51	0.58	0.63	0.64	1.00	0.68
Ensemble	0.83	0.88	0.89	0.91	0.93	0.91	0.90	0.68	1.00

Source: Bloomberg LLC.

correlation is valuable because it introduces more "orthogonal" information that can reduce portfolio variance and enhance risk-adjusted returns, even if the 200d model has a lower standalone Sharpe ratio.

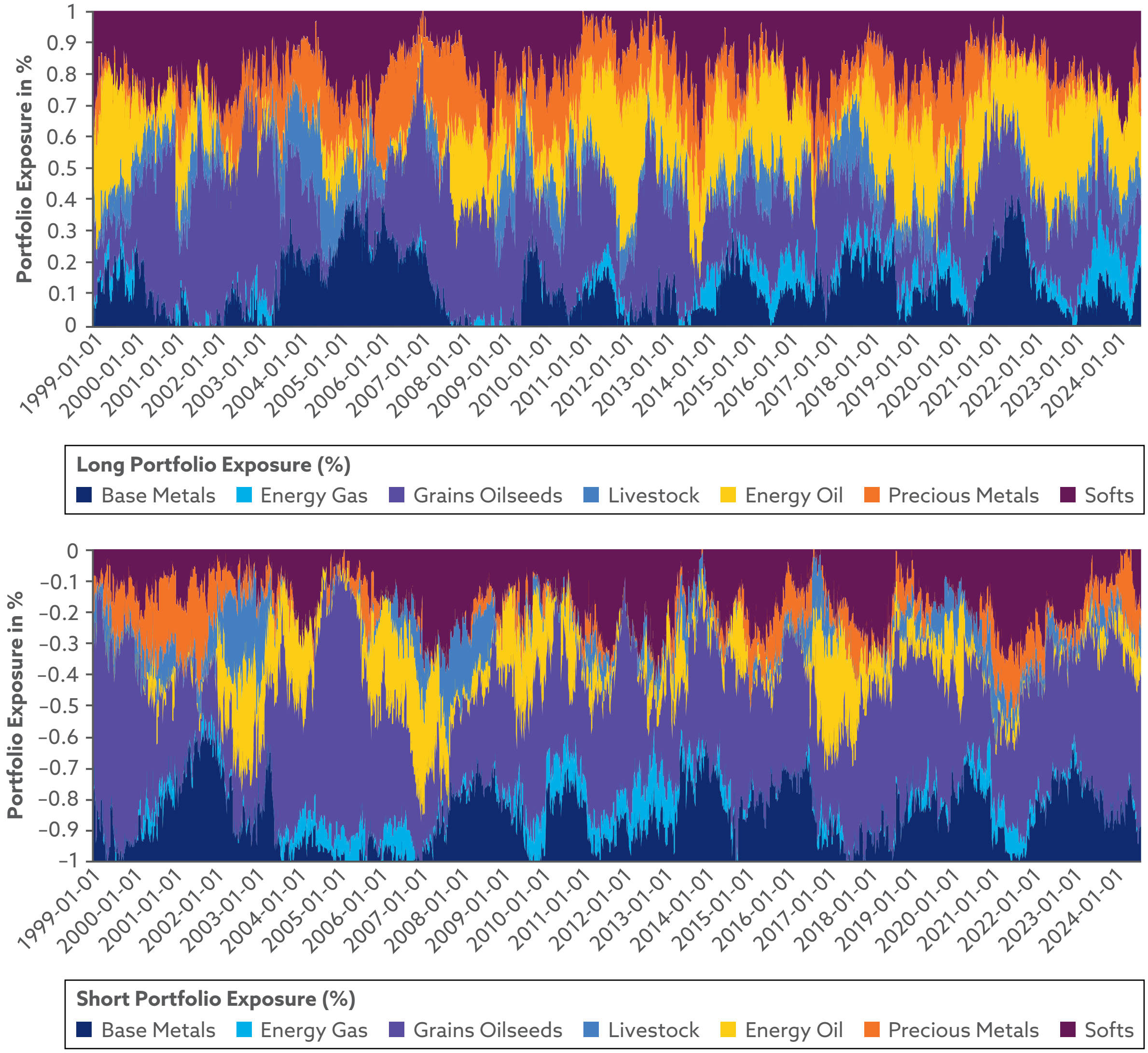
The ensemble model itself exhibits very high correlations with the base strategies, particularly with the mid-term ones, reaching 0.928 with the 66d model and 0.915 with the 88d model. This finding indicates that the ensemble is primarily shaped by these horizons, which may dominate because of superior performance or greater predictive stability. Nevertheless, the ensemble's correlation with the 200d model remains moderate at 0.684, showing that it still benefits from some degree of diversification. The inclusion of all horizons in the ensemble, even those with weaker performance or less commonality, supports the principle of model averaging, where aggregating diverse signals leads to smoother and more robust predictive outputs.

Exhibit 7 shows the long and short book group breakdown of the ensemble portfolio. I classified contracts into subsectors (base metals, precious metals, energy oil, energy gas, softs, grains and oilseeds, livestock) and show the exposure of the long leg and the short leg. Both legs are highly dynamic and invest in all subsectors but with substantial shift in time-varying exposure from net long to net short. When looking at each portfolio separately and focusing on the long leg, grains and oilseeds and precious metals consistently represent a significant portion of the long exposure across the entire sample period. Soft commodities and base metals exhibit elevated exposure during specific intervals, notably between 1999 and 2003, 2009 and 2012, and 2018 and 2021. Crude oil–related positions show pronounced spikes in exposure that coincide with major macroeconomic and geopolitical disruptions, such as those observed in 2008, 2014–2016, and 2022. Natural gas begins to play a more prominent role from 2010 onward, potentially reflecting increased market volatility. Livestock exposures display a more erratic pattern but maintain a persistent presence, with notable peaks occurring in the early 2000s and again in the post-2020 period.

Short exposures are predominantly concentrated in grains and oilseeds, particularly from 1998 to 2011 and after 2022. Natural gas exhibited substantial short allocations between 2006 and 2009 and again from 2019 to 2022, suggesting persistent bearish signals or mean-reverting behavior. Crude oil positions were heavily shorted during 2005–2008 and 2011–2015, coinciding with episodes of elevated volatility and shifting macro conditions. Precious metals and livestock also contribute meaningfully to the short book, with consistent allocations observed throughout 2004–2012. In contrast, base metals and soft commodities exhibit limited or intermittent short exposure, potentially reflecting a structural long bias or weaker signal strength in these sectors.

Finally, **Exhibit 8** shows the net exposure by group. As hinted by the long and short exposure, net market value (NMV) highlights significant temporal variation in directional positions across sectors. Grains and oilseeds show the most pronounced fluctuations, with large positive swings in the early 2000s, 2007–2008, and 2021 and deep negative exposures in 1999–2000, 2005, 2014, and after 2023. Energy oil displays sustained positive net exposure in multiple periods, notably in 2003–2006 and 2012–2015 and from 2020 through 2022. These periods coincide with major energy market cycles, indicating the model's persistent tilt toward oil-related trends during macro-driven dislocations. Livestock maintains consistent, though smaller, net long positions through most of the sample, with occasional reversals during 2008–2010

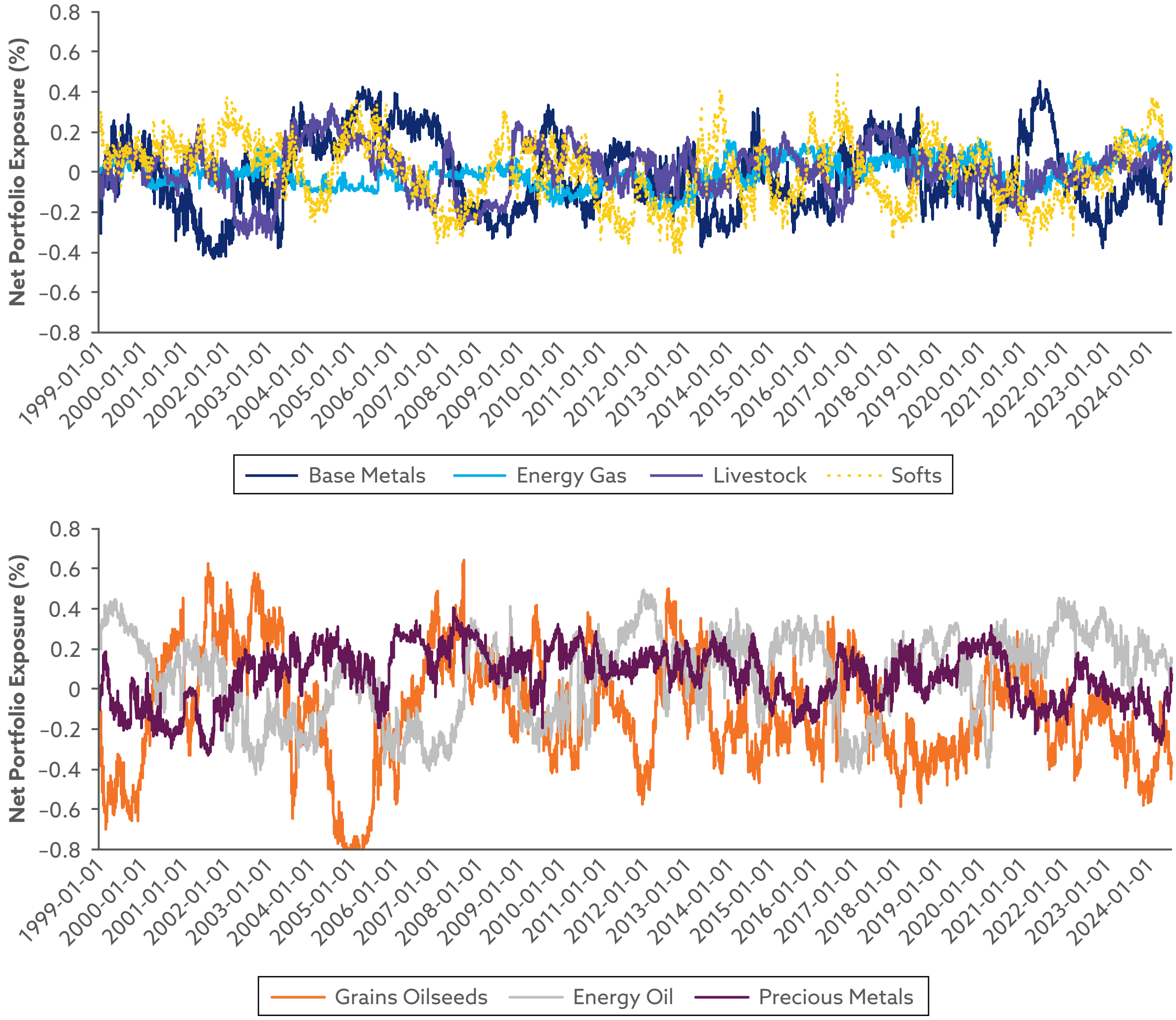
Exhibit 7. Gross Exposure (GMV) Breakdown for Ensemble Long and Short Book



Source: Bloomberg LLC.

and after 2015. Base metals and soft commodities demonstrate alternating positive and negative net exposures, with base metals notably net long during the post–Global Financial Crisis commodity supercycle (2009–2011) and again after the COVID-19 crisis with inflation rising, while softs show long tilts around 2003, 2017–2018, and 2021. Precious metals exhibit notable net short phases in 1999–2000, 2005–2006, and 2013–2015 and after 2022, interspersed with long phases, such as 2008–2009 and the early COVID-19 crisis. These shifts likely reflect both macro-hedging behavior and relative value signals.

Exhibit 8. Net Exposure (NMV) per Contract Type for Ensemble



Source: Bloomberg LLC.

Conclusion

This chapter began with a simple question: Can the rigor and structure of equity-based machine learning pipelines be transposed onto commodity futures markets, where economic intuition runs deep but modeling applications have remained largely siloed and segmented? The results of this study suggest the answer is a clear yes, although not without important structural adaptations relative to equities.

Commodity futures are not stocks. They lack balance sheets, are governed by supply chains rather than corporate disclosures, and move to the beat of exogenous cycles, weather, geopolitics, inventories, and institutional positioning. These characteristics, often framed as modeling challenges, are in fact the very properties that make commodity markets an ideal testbed for modern supervised learning techniques. When organized through a cross-sectional lens, where predictors are derived from asset pricing theory and not just price patterns, commodity futures reveal a persistent and learnable structure that can be used for long–short portfolio construction.

My approach was built on three foundational choices. First, I rooted the features in academic theory, not technical folklore. Carry, basis momentum, skewness, time-series momentum, and so on are not ad hoc constructs but manifestations of storage constraints, hedging pressure, and market segmentation. Second, I extended the empirical asset pricing logic of equity factor models into the commodity space by constructing cross-sectional long–short portfolios guided by machine-learned scores. Finally, I introduced temporal diversity by predicting across multiple horizons, which is a realistic modern way of creating tradable portfolios, capturing both short-term reversals and long-term trend cycles by fusing them via ensemble averaging.

Across a 30-year window and 41 futures contracts, I showed that ML can not only forecast cross-sectional returns but also rank assets in a way that produces consistent economic value. The empirical results are instructive. Time-series momentum, the cornerstone of trend-following CTAs, dominates most models, particularly over intermediate to long horizons. Skewness, a higher-moment feature often ignored in equity ML, plays a meaningful role at short horizons, hinting at reversal tendencies and potential sentiment-driven mispricing. Importantly, the ensemble strategy does more than just blend signals; it reconciles the tension between short-term alpha and long-term robustness. Unlike individual models, which exhibit a mirror J-shaped performance curve across horizons, the ensemble smooths these extremes, offering better drawdown control and volatility reduction without materially sacrificing returns. This finding echoes a broader principle in financial ML: Diversity of signal horizons matters just as much as signal quality.

The correlation structure among horizons further reinforces the value of multiscale modeling. Mid-term models (22d to 100d) cluster tightly, suggesting that they capture variations on a shared theme, likely driven by overlapping features. But the 200-day model, with its low correlations and distinct top features, adds a unique longer-cycle perspective. This decorrelation, although dilutive on a standalone basis, enhances the ensemble's stability and reduces fragility to negative events, as depicted by the better risk profile of the ensemble.

At the portfolio level, the results reveal meaningful heterogeneity. Exposure patterns vary strongly across sectors and time, with pronounced tilts toward grains, energy, and metals at key macro inflection points. These shifts are not noise; they reflect the interaction between predictive signals and regime-specific fundamentals. For instance, long tilts in crude oil during geopolitical crises or short positions in grains after 2022 mirror real-world economic dynamics that models have learned to anticipate.

In conclusion, this chapter bridges the methodological divide between equity ML and commodity return forecasting. It demonstrates that with careful feature construction, robust out-of-sample validation, and thoughtful aggregation across time horizons, ML can serve not only as a modeling tool but as a research framework—one that makes commodities more interpretable, predictable, and, ultimately, investable. As systematic investing continues to evolve, the fusion of domain theory with modern ML holds the key to unlocking latent structure in every markets.

References

Ampountolas, Apostolos, and Sayed AlGharbi. 2025. "An Innovative Hybrid LightGBM-BPNN Model for Enhanced Commodity Forecasting Accuracy." *Finance Research Open* 1 (1). doi:10.1016/j.finr.2025.100004.

Anderson, Bing, and Shuyun May Li. 2015. "An Investigation of the Relative Strength Index." *Banks & Bank Systems* 10 (1): 92–96.

Angelidis, Timotheos, Athanasios Sakkas, and Nikalaos Tessaromatis. 2025. "Predicting Commodity Returns: Time Series vs. Cross Sectional Prediction Models." *Journal of Commodity Markets* 38 (June). doi:10.1016/j.jcomm.2025.100475.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen. 2013. "Value and Momentum Everywhere." *Journal of Finance* 68 (3): 929–85. doi:10.1111/jofi.12021.

Basu, Devraj, and Joëlle Miffre. 2013. "Capturing the Risk Premium of Commodity Futures: The Role of Hedging Pressure." *Journal of Banking & Finance* 37 (7): 2652–64. doi:10.1016/j.jbankfin.2013.02.031.

Ben Jabeur, Sami, Rabeh Khalfaoui, and Wissal Ben Arfi. 2021. "The Effect of Green Energy, Global Environmental Indexes, and Stock Markets in Predicting Oil Price Crashes: Evidence from Explainable Machine Learning." *Journal of Environmental Management* 298 (November). doi:10.1016/j.jenvman.2021.113511.

Bhardwaj, Geetesh, Gary Gorton, and Geert Rouwenhorst. 2015. "Facts and Fantasies About Commodity Futures Ten Years Later." NBER Working Paper 21243 (June). doi:10.3386/w21243.

Blitz, David, Matthias X. Hanauer, Tobias Hoogteijling, and Clint Howard. 2023. "The Term Structure of Machine Learning Alpha." Working paper (19 July). doi:10.2139/ssrn.4474637.

Boons, Martijn, and Melissa Porras Prado. 2019. "Basis-Momentum." *Journal of Finance* 74 (1): 239–79. doi:10.1111/jofi.12738.

Cheng, Ing-Haw, Andrei Kirilenko, and Wei Xiong. 2015. "Convective Risk Flows in Commodity Futures Markets." *Review of Finance* 19 (5): 1733–81. doi:10.1093/rof/rfu043.

Costa, Alexandre Bonnet R., Pedro Calvacanti G. Ferreira, Wagner P. Gaglianone, Osmani Teixeira C. Guillén, João Victor Issler, and Yihao Lin. 2021. "Machine Learning and Oil Price Point and Density Forecasting." *Energy Economics* 102 (October). doi:10.1016/j.eneco.2021.105494.

de Roon, Frans A., Theo E. Nijman, and Chris Veld. 2000. "Hedging Pressure Effects in Futures Markets." *Journal of Finance* 55 (3): 1437–56. doi:10.1111/0022-1082.00253.

Fama, Eugene F., and Kenneth R. French. 2015. "Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage." In *The World Scientific Handbook of Futures Markets*, edited by Anastasio G. Malliaris and William T. Ziemba, 79–102. London: World Scientific.

Fernandez-Perez, Adrian, Bart Frijns, Ana-Maria Fuertes, and Joëlle Miffre. 2018. "The Skewness of Commodity Futures Returns." *Journal of Banking & Finance* 86 (January): 143–58. doi:10.1016/j.jbankfin.2017.06.015.

Foroutan, Parisa, and Salim Lahmiri. 2024. “Deep Learning Systems for Forecasting the Prices of Crude Oil and Precious Metals.” *Financial Innovation* 10 (16 July). doi:10.1186/s40854-024-00637-z.

Fuertes, Ana-Maria, Joëlle Miffre, and Adrian Fernandez-Perez. 2015. “Commodity Strategies Based on Momentum, Term Structure, and Idiosyncratic Volatility.” *Journal of Futures Markets* 35 (3): 274–97. doi:10.1002/fut.21656.

Gargano, Antonio, and Allan Timmermann. 2014. “Forecasting Commodity Price Indexes Using Macroeconomic and Financial Predictors.” *International Journal of Forecasting* 30 (3): 825–43. doi:10.1016/j.ijforecast.2013.09.003.

Gorton, Gary B., Fumio Hayashi, and K. Geert Rouwenhorst. 2013. “The Fundamentals of Commodity Futures Returns.” *Review of Finance* 17 (1): 35–105. doi:10.1093/rof/rfs019.

Gorton, Gary, and K. Geert Rouwenhorst. 2006. “Facts and Fantasies About Commodity Futures.” *Financial Analysts Journal* 62 (2): 47–68. doi:10.2469/faj.v62.n2.4083.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. “Empirical Asset Pricing via Machine Learning.” *Review of Financial Studies* 33 (5): 2223–73. doi:10.1093/rfs/hhaa009.

Guida, Tony, and Guillaume Coqueret. 2019. “Ensemble Learning Applied to Quant Equity: Gradient Boosting in a Multifactor Framework.” In *Big Data and Machine Learning in Quantitative Investment*, edited by Tony Guida, 129–148. Hoboken, NJ: Wiley.

Hirshleifer, David. 1988. “Residual Risk, Trading Costs, and Commodity Futures Risk Premia.” *Review of Financial Studies* 1 (2): 173–93. doi:10.1093/rfs/1.2.173.

Hong, Harrison, and Motohiro Yogo. 2012. “What Does Futures Market Interest Tell Us About the Macroeconomy and Asset Prices?” *Journal of Financial Economics* 105 (3): 473–90. doi:10.1016/j.jfineco.2012.04.005.

Jegadeesh, Narasimhan, and Sheridan Titman. 2001. “Profitability of Momentum Strategies: An Evaluation of Alternative Explanations.” *Journal of Finance* 56 (2): 699–720. doi:10.1111/0022-1082.00342.

Kaldor, Nicholas. 1939. “Speculation and Economic Stability.” *Review of Economic Studies* 7 (1): 1–27. doi:10.2307/2967593.

Keynes, John M. 1930. *A Treatise on Money, Volume 1: The Pure Theory of Money*. London: Harcourt, Brace & Co. doi:10.4324/9781003547150-7.

Lucas, Robert E., Jr. 1978. “Asset Prices in an Exchange Economy.” *Econometrica* 46 (6): 1429–45. doi:10.2307/1913837.

Lukac, Louis P., B. Wade Brorsen, and Scott H. Irwin. 1988. “A Test of Futures Market Disequilibrium Using Twelve Different Technical Trading Systems.” *Applied Economics* 20 (5): 623–39. doi:10.1080/00036848800000113.

Miffre, Joëlle, and Georgios Rallis. 2007. “Momentum Strategies in Commodity Futures Markets.” *Journal of Banking & Finance* 31 (6): 1863–86. doi:10.1016/j.jbankfin.2006.12.005.

Moskowitz, Tobias J., Yao Hua Ooi, and Lasse Heje Pedersen. 2012. “Time Series Momentum.” *Journal of Financial Economics* 104 (2): 228–50. doi:10.1016/j.jfineco.2011.11.003.

Narayan, Paresh Kumar, Hudson Ali Ahmed, and Seema Narayan. 2015. “Do Momentum-Based Trading Strategies Work in the Commodity Futures Markets?” *Journal of Futures Markets* 35 (9): 868–91. doi:10.1002/fut.21685.

Park, Cheol-Ho, and Scott H. Irwin. 2007. “What Do We Know About the Profitability of Technical Analysis?” *Journal of Economic Surveys* 21 (4): 786–826. doi:10.1111/j.1467-6419.2007.00519.x.

Paschke, Raphael, Marcel Prokopczuk, and Chardin Wese Simen. 2020. “Curve Momentum.” *Journal of Banking & Finance* 113 (April). doi:10.1016/j.jbankfin.2019.105718.

Shapley, Lloyd S. 1953. “A Value for *n*-Person Games.” In *Contributions to the Theory of Games II*, edited by Harold W. Kuhn and Albert William Tucker, 307–17. Princeton, NJ: Princeton University Press.

Simonian, Joseph, Chenwei Wu, Daniel Itano, and Vyshaal Narayanam. 2019. “A Machine Learning Approach to Risk Factors: A Case Study Using the Fama–French–Carhart Model.” *Journal of Financial Data Science* 1 (1): 32–44. doi:10.3905/jfds.2019.1.032.

Szakmary, Andrew C., Qian Shen, and Subhash C. Sharma. 2010. “Trend-Following Trading Strategies in Commodity Futures: A Re-Examination.” *Journal of Banking & Finance* 34 (2): 409–26. doi:10.1016/j.jbankfin.2009.08.004.

Wang, Shirui, and Tianyang Zhang. 2024. “Predictability of Commodity Futures Returns with Machine Learning Models.” *Journal of Futures Markets* 44 (2): 302–22. doi:10.1002/fut.22471.

Working, Holbrook. 1949. “The Theory of Price of Storage.” *American Economic Review* 39 (6): 1254–62. www.jstor.org/stable/1816601.

Yen, Stéphane Meng-Feng, and Ying-Lin Hsu. 2010. “Profitability of Technical Analysis in Financial and Commodity Futures Markets—A Reality Check.” *Decision Support Systems* 50 (1): 128–39. doi:10.1016/j.dss.2010.07.008.