Final Project- Refining Equities Analysis- Return and Volatility Modeling

MSDS 492- Analysis of Financial Markets

Northwestern University

1/30/2026

By David Van Dyke

AI used in this report for custom coding, background search, and wording clarification. The work is my own.

Contents

Abstract

This study investigates the daily stock return behavior of three major U.S. petroleum refining companies. The review covers the companies Phillips 66 (PSX), Valero Energy Corporation (VLO), and Marathon Petroleum Corporation (MPC) over a 15-year period to understand how market and commodity conditions influence short-term price movements. The research is conducted in two phases. The first phase provides a comparative analysis of return distributions, volatility patterns, and sensitivity to key external drivers, including the S&P 500 Index, crude oil prices, gold prices, and market volatility indicators(VIX). These diagnostics identify similarities and differences among the three companies and guide the selection of one firm for focused modeling.

In the second phase, the study develops a simplified predictive framework aimed at forecasting the binary direction of next-day returns (up or down) for the selected company. A streamlined set of predictors from the equity market returns, commodity price changes, and shifts in market volatility will be used to build and compare three modeling approaches: a baseline logistic regression model, a GARCH-based volatility-informed

signal, and a machine-learning classification method. Prediction accuracy metrics are used to evaluate model performance. Together, the results provide practical insights into the directional predictability of daily stock movements within the refining sector and demonstrate an easier to understand methodology for short-horizon forecasting.

Keywords

Phillips 66, Valero Energy, Marathon Petroleum, daily returns, stock volatility, binary prediction, directional forecasting, logistic regression, GARCH, machine learning, refining sector

Introduction

The petroleum refining and downstream energy sector faces constant exposure to volatile commodity markets, shifting macroeconomic conditions, and unpredictable equity market behavior. Companies such as Phillips 66 (PSX), Valero Energy Corporation (VLO), and Marathon Petroleum Corporation (MPC) exhibit daily stock movements that are particularly sensitive to fluctuations in crude oil prices, broader market returns, and changes in investor risk sentiment. Despite the importance of short-term price behavior, predicting the daily direction of stock returns is a difficult challenge. This is largely due to the rapid and often nonlinear responses of these stocks to external forces similar to the Efficient Market Hypothesis predictions (EMH). This study is motivated by the need to better understand these short-horizon dynamics and to address a practical problem: organizations operating in or adjacent to the energy sector frequently require timely insights for hedging decisions, risk management, liquidity planning, and tactical market positioning. However, there is limited accessible research that offers a clear and manageable framework for forecasting whether a stock will move up or down the next day.

The study addresses this gap through a two-phase approach. First, it analyzes the daily return and volatility behavior of PSX, VLO, and MPC over a 15-year period, comparing their distributional characteristics, market correlations, and sensitivities to key financial and commodity indicators. The purpose of this phase is to understand the general nature of daily stock behavior across the refining sector and to determine which company presents the most promising structure for predictive modeling. The second phase focuses on developing a simplified forecasting model for one selected company, using market and commodity variables such as S&P 500 returns, crude oil and gold price movements, and changes in market volatility (VIX) as predictors. The central research questions are: How do daily return patterns differ across PSX, VLO, and MPC? Which external variables are most closely associated with next-day directional shifts? And can a streamlined predictive framework accurately classify whether the stock will move up or down tomorrow? Based on existing financial theory and sector behavior, the study hypothesizes that daily direction is materially influenced by broad market returns, oil price changes, and volatility conditions, and that a simplified classification model can outperform random chance.

This problem is meaningful not only in an academic sense but also in an organizational context. Firms involved in trading, investment management, risk oversight, and operational planning often rely on short-term forecasts to respond to market shocks and shifting conditions. Improved understanding of directional movements can support more effective hedging strategies, short-term financial decision-making, and enhanced anticipation of price risks. By combining cross-company diagnostics with a focused predictive model, this study offers an approachable yet informative framework for analyzing and forecasting daily stock behavior.

Literature Review

Predicting short-horizon stock returns, especially for energy-linked equities, remains an active research area. Recent studies increasingly move beyond traditional econometric models toward hybrid GARCH–machine-learning frameworks and nonlinear predictive architectures to better capture volatility clustering, spillovers, and commodity–equity interactions. This study aligns with these concepts but differs in its focus on daily directional forecasting within U.S. refining companies.

1. Volatility Modeling and GARCH (Generalized Autoregressive Conditional Heteroskedasticity)

Recent work refines GARCH-based approaches through hybrid and multicomponent models. Studies comparing univariate and multivariate GARCH with machine-learning predictors show that ML models often outperform GARCH in out-of-sample volatility forecasting, though each exhibits systematic bias in opposite directions.

Other research integrates macroeconomic and policy-uncertainty variables using GARCH-MIDAS(Mixed Data Sampling), finding that long-term volatility components across commodities are shaped by economic activity, sentiment, and policy uncertainty.

These efforts share the study's concern with short-horizon predictability, but differ in modeling volatility levels directly. This study instead uses volatility indicators (e.g., VIX changes) as predictors in a directional classification setting, reducing model complexity while retaining informational value.

2. Machine-Learning Methods for Return Prediction

Deep learning and ensemble methods are increasingly applied to short-term return forecasting. Neural networks using firm-level and macroeconomic inputs improve predictive performance by capturing nonlinear interactions.

These related studies address similar decision problems for how to outperform naïve baselines in noisy financial settings. However, the other studies more complex approach contrasts with this streamlined design aimed at operational usability.

3. Commodity–Equity Spillovers and Cross-Market Dynamics

Because refining firms are highly exposed to energy inputs, commodity–equity linkages are central. Research documents strong volatility spillovers among major commodities and between oil and equity markets.

Additional work shows time-varying relationships between stock-market volatility and commodity prices, with patterns differing across crisis periods. Cross-market spillover analyses reveal that significant return-forecast error variance arises from shocks transmitted across commodities and equity markets.

This study incorporates similar cross-market drivers. The other drivers include S&P 500 returns, crude-oil price changes, gold price changes, and changes in volatility indices(VIX). This diverges from the other studies by applying the drivers to binary direction forecasting rather than modeling spillover magnitudes.

4. Energy-Sector Return Behavior and Sector-Specific Predictability

Broader oil-and-gas equity research shows refining and marketing firms tend to exhibit high daily expected returns and elevated sensitivity to equity-market risk. Sector-level studies find oil-price predictability varies by industry characteristics such as size, book-to-market ratio, and trading volume. The recent studies also highlights the role of energy-related information, including news narratives, in shaping return predictability.

These studies frame sector impacts but do not examine downstream refiners specifically or focus on daily directional outcomes. This study narrows this gap by building a simple, interpretable, and sector-specific framework tailored to U.S. refining stocks.

Volatility Modeling & GARCH / Hybrid Models

Chung (2024) – Modelling and forecasting energy market volatility using GARCH and machine learning

https://arxiv.org/html/2405.19849v1

Lasisi et al. (2025) – Commodity market volatility with climate policy uncertainty (GARCH–MIDAS)

https://link.springer.com/article/10.1007/s43546-025-00792-0

Machine Learning & Deep Learning for Stock Prediction

Wang (2024) – Stock return prediction with neural network models

https://link.springer.com/article/10.1186/s40854-023-00608-w

Liu (2025) – Advancing Stock Return Prediction:: Comprehensive study of traditional ML and deep learning models

https://www.atlantis-press.com/article/126008539.pdf

Peng (2024) – Comparative evaluation of RF, SVM, BPNN for stock prediction

https://dl.acm.org/doi/full/10.1145/3705618.3705665

AI-Enhanced Hybrid Prediction Approaches

Cohen et al. (2025) – Hybrid ML + LLM predictive framework (Entropy)

https://www.mdpi.com/1099-4300/27/6/550

Energy Information & Return Predictability

Zhang & Wang (2025) – Energy information and stock return predictability

https://link.springer.com/content/pdf/10.1007/978-981-96-7655-2_9.pdf?pdf=inline%20link

Degiannakis et al. (2018) – Oil prices & stock market relationships

https://www.jstor.org/stable/26534476

Oil Price Forecasting & Multi-Model Studies

Cohen (2025) – Short-term oil price forecasting using econometric & ML techniques

https://www.mdpi.com/2504-4990/7/4/127

Bashiri Behmiri & Manso (2021) – Comprehensive review of oil forecasting techniques

https://1library.net/document/y876375z-crude-oil-price-forecasting-techniques-comprehensive-review-literature.html

Commodity–Equity Spillovers & Volatility Dynamics

Melas et al. (2024) – Volatility spillovers among major commodities

https://www.mdpi.com/1911-8074/17/8/365

Kang, Perez de Gracia & Ratti (2025) – Stock market volatility & commodity prices

https://www.cambridge.org/core/journals/macroeconomic-dynamics/article/stock-market-volatility-and-commodity-prices/28B4C94D275EEF206FAC7F0C4B493F32

Pinho & Maldonado (2022) – Commodity & equity market spillovers

https://www.researchgate.net/publication/367603538_Commodity_and_Equity_Markets_Volatility_and_Return_Spillovers/fulltext/63d94523c465a873a271ebff/367603538_Commodity_and_Equity_Markets_Volatility_and_Return_Spillovers.pdf

Long-Term Commodity Volatility Drivers

Nguyen & Walther (2017) – Long-term volatility drivers of commodity markets (GARCH-MIDAS)

https://mpra.ub.uni-muenchen.de/84464/1/MPRA_paper_84464.pdf

Sector-Level Equity Return Research

Carson (2022) – Long-term daily equity returns across oil & gas sectors

https://link.springer.com/article/10.1007/s10842-021-00374-4

Phan, Sharma & Narayan (2015) – Stock return forecasting: new evidence

https://research.monash.edu/en/publications/stock-return-forecasting-some-new-evidence

Theoretical Framework

The theoretical foundation of this study rests on several interconnected principles from financial economics that explain why daily stock movements in the U.S. refining sector may display short-horizon predictability. A natural starting point is the Efficient Market Hypothesis (EMH), which posits that asset prices embed all available information and therefore follow a largely unpredictable path. Yet empirical research has long shown that even in efficient markets, short-term patterns such as return autocorrelation, reversals, and volatility clustering can emerge—especially in industries exposed to rapid information flow and commodity-driven shocks. Refining companies like PSX, MPC, and VLO are tightly linked to global energy markets, making them sensitive to shifts in crude oil prices, broad equity-market performance, and sudden changes in investor risk sentiment. These mechanisms create the conditions under which the next-day direction of stock returns may reflect systematic responses to external information rather than pure randomness.

A second core theoretical pillar involves volatility dynamics, particularly the well-documented phenomenon of volatility clustering. Financial markets, and energy equities in particular, frequently experience periods of heightened turbulence followed by additional large movements. GARCH-type theories formalize this behavior by modeling conditional variance as dependent on past shocks. Although this study does not estimate full GARCH models, it draws from their conceptual insights that namely, the volatility today provides information about uncertainty tomorrow. Indicators such as the VIX serve as practical proxies for market stress, capturing investors' evolving risk perceptions. Incorporating changes in VIX into next-day forecasts therefore aligns with theoretical expectations about volatility persistence and its influence on trading behavior.

These ideas connect naturally to the broader literature on cross-market spillovers and transmission effects. Refining companies operate at the nexus of commodity and equity markets, meaning their stock prices respond not only to firm-specific conditions but also to movements in crude oil, gold, and macroeconomic indicators. Crude oil returns capture changes in input costs and refining margins, while gold prices reflect flight-to-quality dynamics that often accompany shifts in economic sentiment. Similarly, S&P 500 returns represent systematic market forces that shape sector-level exposure to broad economic cycles. These interdependent relationships are consistent with spillover theory, which suggests that shocks in one market can influence price behavior in another, creating a foundation for short-term directional predictability even when exact long-run forecasting remains limited.

Lastly, the theoretical motivations for the modeling approaches used in this study stem from how different predictive methods handle signal and noise in financial data. Logistic regression offers an interpretable baseline for estimating directional probabilities based on linear relationships in the predictors. Volatility-aware signals inspired by GARCH theory allow the model to account for shifts in market uncertainty without the complexity of full volatility modeling. Machine-learning classifiers, by contrast, provide flexibility to capture nonlinear interactions and threshold effects that traditional models may overlook. Together, these approaches form a cohesive predictive framework grounded in the realities of high-noise financial environments, allowing the study to balance interpretability with the ability to detect subtle patterns across equity, commodity, and volatility markets.

Data and Methodology

This study relies on daily financial market data sourced exclusively from Yahoo Finance, which provides publicly available historical pricing for equities, commodities, and market indices. All data used in the analysis reflects daily adjusted closing prices to ensure that dividends, stock splits, and contract-roll adjustments (for futures) are properly incorporated. The dataset spans a 15-year period and includes seven key tickers selected to represent refining-sector equities, benchmark equity-market performance, commodity inputs, and broad measures of market volatility. Each ticker plays a distinct conceptual role in the modeling framework.

The three primary equity tickers analyzed are MPC (Marathon Petroleum Corporation), PSX (Phillips 66), and VLO (Valero Energy Corporation). These firms represent the major publicly traded U.S. downstream refining companies. Marathon Petroleum (MPC) is one of the largest independent refiners in North America, operating extensive refining, marketing, and midstream networks. Phillips 66 (PSX) is a diversified downstream and midstream company with refining, chemical, and energy-transportation assets, making it highly sensitive to both refining margins and broader energy-market conditions. Valero Energy (VLO) is a leading independent refiner with a global refining footprint and a business model closely tied to crude-oil processing economics. Together, these companies provide a comprehensive view of refining-sector stock behavior and offer a basis for comparative analysis and predictive modeling.

To capture systematic market influences, the study includes ^GSPC, which represents the S&P 500 Index. This index is a widely used benchmark for overall U.S. equity-market conditions and reflects broad macroeconomic sentiment, risk valuation, and investor positioning. Movements in the S&P 500 provide essential context for interpreting whether refining-sector price changes are driven by firm-specific dynamics or by market-wide forces that influence nearly all equities.

Because refiners are deeply linked to commodity markets, two crucial commodity tickers are included: CL=F (Crude Oil Futures, WTI) and GC=F (Gold Futures). Crude-oil prices represent the primary input cost for refining companies and strongly influence refining margins, profitability expectations, and sector-level risk pricing. Gold, while not directly tied to the refining value chain, acts as a global safe-haven asset. Changes in gold prices often reflect shifts in investor risk aversion, inflation expectations, and macroeconomic uncertainty which are factors that can indirectly influence equity markets, including energy stocks. Incorporating both energy-linked and sentiment-sensitive commodities allows the analysis to capture a wider spectrum of cross-market interactions.

Finally, the dataset includes ^VIX, the CBOE Volatility Index, which measures implied volatility in S&P 500 index options. Often referred to as the “fear index,” the VIX reflects market expectations of near-term volatility and serves as a high-frequency indicator of investor uncertainty. For daily prediction models, changes in VIX can provide meaningful information about evolving risk sentiment that may influence next-day directional movements in energy equities.

Although the research is designed around a nominal 15-year horizon, not all instruments have uninterrupted trading histories spanning that full period. Specifically, Phillips 66 (PSX) and Marathon Petroleum Corporation (MPC) began trading in 2012 and 2011, respectively, following corporate spin-offs, and therefore their available price histories fall slightly short of the full 15-year window. All other tickers in the dataset Valero Energy (VLO), S&P 500 Index (^GSPC), WTI Crude Oil Futures (CL=F), Gold Futures (GC=F), and the VIX Index

(^VIX) have substantially longer trading histories and fully cover the intended analysis horizon.

From these seven tickers, daily log returns were computed to standardize comparisons across markets and reduce the influence of price-level differences. Exploratory diagnostics including distributional analysis, autocorrelation structure, volatility persistence, and cross-asset return correlations were conducted to guide model design and justify variable selection. Building on these diagnostics, the study implemented three predictive approaches: a baseline logistic regression model using daily predictors, a volatility-aware model incorporating VIX dynamics, and a machine-learning classifier designed to capture potential nonlinear relationships. Each model uses the same set of daily return-based predictors derived from the Yahoo Finance data, ensuring consistency and comparability across the modeling phase.

Engineered Features for Nonlinear Signals

To complement the baseline feature set of daily market and commodity returns, the study incorporates a structured set of engineered features designed to capture nonlinear behaviors and cross-market interactions that a simple logistic specification may otherwise miss. These augmented inputs draw on both financial-economics intuition and sector-specific dynamics relevant to refining-equity behavior. The goal is not to increase complexity arbitrarily, but to introduce features that have a clear economic interpretation and the potential to meaningfully improve directional classification accuracy.

First, we experiment with interaction-based composite indicators that encode scenarios where simultaneous movements across markets convey information beyond the sum of their individual effects. For instance, a binary “Equity Down & Oil Down” flag is constructed to signal joint weakness in the broad equity market and the crude complex, a combination that may exert disproportionate downside pressure on refining names relative to a linear model of S&P 500 and oil effects alone. Similarly, a “Risk-Off Day” indicator—defined by the concurrent pattern of Gold rising, the S&P 500 declining, and the VIX increasing—captures episodes of broad risk aversion that may lead to accelerated defensive rotation and short-term underperformance among refiner equities.

Next, we introduce regime-classification features to represent structural shifts in volatility or market stress. A high-volatility dummy is activated when realized or implied volatility surpasses a threshold consistent with a historically elevated risk environment. In addition, a crude-oil shock flag is triggered when the daily change in WTI exceeds two standard deviations of its rolling historical distribution. Because refiner equity performance can respond disproportionately to large commodity shocks such as margin compression from sharp crude rallies these indicators allow the model to adapt its behavior across market regimes rather than relying on a single linear relationship.

To allow the model to describe short-term return dynamics more flexibly, we incorporate lagged return features beyond the standard one-day lag. If initial model diagnostics reveal systematic patterns such as intraday momentum or two-day reversals then including a two-day-lagged return can help capture short bursts of trend persistence or correction that influence directional movement on extremely short horizons.

We further include features designed to measure relative sector performance. A simple “excess return” metric, defined as the refiner’s daily return minus the S&P 500 return, serves as a compact way to identify whether the stock is outperforming or lagging the broader market at a given moment. Periods of notable underperformance may signal subsequent mean reversion, while episodes of unusual strength may fade, depending on prevailing sector conditions.

Finally, we add a refining-margin proxy, constructed as a spread between crude and equity-based indicators such as Oil minus the S&P 500 return, or Oil minus the average return of the refiner group. Because refining margins often behave inversely to crude prices, and because margin expectations play a central role in refiner equity valuation, these spreads offer low-complexity approximations of the crack-spread sentiment that can influence next-day price direction.

Each engineered feature is evaluated iteratively using cross-validation, with retention based solely on out-of-sample performance improvements. By adhering to the principle of economic plausibility, the study avoids unnecessary model opacity and maintains interpretability while enabling the logistic and machine-learning models to capture richer, sector-specific patterns in daily directional movements.

Economic Value Evaluation of Predictive Signals

Beyond statistical accuracy reviewed in this study, future work could be done on evaluating the economic value of the forecasting framework by simulating a simple, transparent trading strategy built directly on the model’s daily predictions. Traditional accuracy metrics alone do not capture whether predicted signals translate into improved financial outcomes, particularly in high-noise equity environments where even modest forecast skill may generate meaningful returns if applied systematically. The economic-value analysis therefore serves as a practical test of whether the model’s insights hold operational and investment relevance.

The simulated strategy operates on a one-day horizon. Each trading day in the test set, the model's output determines the portfolio position: if the model predicts an "Up" day for VLO, the strategy takes a long position for the following session; if a "Down" day is predicted, the strategy switches into a short position, depending on the specification being tested. Some type of loss protection would also be incorporated. At day's end, the position is closed, and the process repeats. This structure directly links predictive skill to realized return outcomes, without requiring leverage, derivatives, or complex rebalancing rules.

To benchmark performance, the cumulative returns of this model-driven strategy are compared to two baselines:

- Buy-and-hold, which remains fully invested at all times, and
- A random 50/50 strategy, which mirrors the same daily trading mechanics but selects positions randomly rather than based on predictions.

This comparison could potentially assess whether the model offers value above passive exposure and beyond what could be achieved by chance.

Performance could be evaluated using standard portfolio metrics, including annualized return, annualized volatility, Sharpe ratio, and maximum drawdown. These measures capture not only raw return generation but also the risk profile and stability of the strategy. A model that offers materially higher risk-adjusted returns or a smaller drawdown than buy-and-hold demonstrates meaningful predictive value even if its accuracy advantage is modest in absolute terms.

To ensure realism, the analysis would incorporate transaction cost assumptions, such as a 0.1% cost per trade. Because short-horizon strategies can be vulnerable to turnover-driven slippage, using conservative cost estimates acts as an important constraint: only predictive signals that deliver economically robust alpha after friction are considered meaningful. This prevents overstating the value of a model that predicts direction successfully but requires overly frequent trading to realize those gains.

Ultimately, demonstrating that the model produces positive alpha, higher Sharpe ratios, or reduced downside risk relative to passive benchmarks underscores the real-world significance of the predictive framework. The economic-value evaluation therefore serves as the final and most stringent test of model usefulness, complementing traditional statistical validation and ensuring that the study's conclusions emphasize actionable, not merely theoretical, predictive power.

Results

Initial ideas on model evaluation

Model Validation and Performance Metrics

We will evaluate model performance comprehensively using classification metrics:

Classification Metrics: Primary metric is Directional Accuracy i.e. the percentage of days the model correctly predicts Up vs Down. This will be compared against the 50% benchmark (which is the expected accuracy of random guessing or an efficient market with no predictability). We will also compute the True Positive Rate (accuracy on actual Up days) and True Negative Rate (accuracy on Down days) to see if the model is better at calling ups vs downs. Considering the class balance: in our 15-year sample, roughly 50% of days are up and 50% down (which is typical for stock returns), so accuracy is a suitable metric. We will nevertheless look at the Brier score or log-loss for probabilistic calibration if using logistic/ML probabilities to assess if the predicted probabilities (e.g., 60% chance up) are well calibrated. Additionally, Area Under ROC Curve (AUC) may be reported to summarize overall classification performance independent of threshold. However, since we ultimately act on a hard binary decision, accuracy and error rates are most interpretable. A confusion matrix will be presented to show counts of True Ups, False Ups, etc., for each model.

Statistical Significance: Using the bootstrap of test results, we will derive a confidence interval for accuracy. For example, if the logistic model is correct on 55% of ~750 test days, we'll check if 0.55 is significantly above 0.50 (it likely would be, with $p<0.05$, since at 750 samples the standard error for 50% accuracy is ~1.8%). We may also apply the Henriksson-Merton test or Pesaran-Timmermann test – statistical tests used in finance for market-timing ability – to confirm that the directional forecasts add value beyond chance. Significance gives credence to the results as a genuine predictive signal rather than luck.

Comparison across Models: We will tabulate the accuracy (and other metrics) of the three approaches side by side. This will show, for instance, if the Machine Learning classifier outperforms logistic by a few percentage points, or if logistic is essentially as good. If the GARCH-based approach is implemented as a standalone classifier (perhaps a simple rule-based classifier using volatility), we'll include its performance as well. We anticipate something like: logistic accuracy ~55%, RF accuracy maybe 57%, etc., but these are to be validated. If differences are small, we will favor the simpler model (logistic) for parsimony – unless the ML's edge is consistent and significant.

Volatility Prediction Metrics: Although our primary goal is direction, the GARCH model's own performance on predicting volatility will be evaluated with standard metrics (for completeness). We'll compute the MSE between predicted volatility (or standard deviation) and the realized volatility (e.g., square of next day's return) This will tell us how well the GARCH is doing at forecasting magnitude. If the volatility forecast is poor, its value in direction prediction might be limited. Conversely, if volatility forecasts are good, we might integrate them more

Conclusions