Working Paper: Topological Regimes in Commodity Markets: Detecting Loop Structures via Persistent Homology

Anirudh Krovi[1]

[1]PhD, Northwestern University; MBA, NYU Stern; Formerly at McKinsey & Company,
anirudh.krovi@stern.nyu.edu

July 11, 2025

Abstract

We propose a novel framework for detecting regime structure in commodity markets using persistent homology. Applying this method to crude oil and wheat price trajectories, we extract sliding windows of 10-day averaged prices and project them into low-dimensional space via principal component analysis (PCA). We then compute Betti-1 persistence to identify loop-like structures in these trajectories, which we interpret as indicators of regime cyclicity. Our analysis reveals that loop intensity varies across commodities and is predictive of future behavior. By combining topological features with standard statistical descriptors, we build simple yet effective classifiers and backtest trading strategies with favorable risk-adjusted returns. These results suggest that topological signals offer a compact and interpretable alternative to classical regime modeling techniques.

1 Introduction

Topological techniques have emerged as powerful tools for data analysis and machine learning, offering a geometric lens on complex datasets. Unlike traditional statistical methods that focus on local patterns or learned embeddings, topological data analysis (TDA) captures global structures — such as connectivity, cycles, and voids — that persist across scales. These features have proven valuable in diverse domains including image analysis, neuroscience, and genomics [3, 4, 13].

A central tool in TDA is *persistent homology*, which quantifies the presence of topological features over multiple resolutions. Given a point cloud or time series, persistent homology extracts a *persistence diagram* summarizing the birth and death of features like connected components (Betti-0), loops (Betti-1), and voids (Betti-2) [9, 5]. This approach has been used to identify periodicity in dynamical systems [16], structural motifs in time series [21], and latent geometry in embedding spaces [17]. Despite its successes, persistent homology remains underutilized in finance — particularly in modeling price trajectories.

Commodity markets present an especially compelling use case. Prices in these markets often reflect seasonal cycles, supply-demand feedback loops, or abrupt regime shifts due to geopolitical or macroeconomic shocks. Traditional approaches — including ARIMA, GARCH, Fourier analysis, and regime-switching models like hidden Markov models (HMMs) — primarily capture trends, volatility, and latent states via scalar features [12, 1, 6]. Recent work in machine learning has introduced neural networks and decision tree ensembles for commodity forecasting [11, 20], but such models still overlook the shape of the price trajectory itself.

In this paper, we introduce a topological framework for commodity price modeling that uses persistent homology to detect regime-like behavior via loop structures in price trajectories. Specifically, we extract 10-day averaged price windows from daily data, project them into low-dimensional space via PCA, and compute Betti-1 — a measure of loopiness — for each window. We interpret strong Betti-1 signals as indicators of local cyclicity or structured movement, and weak signals as reflecting trendless or shock-driven behavior.

Importantly, we do not assume all commodities behave alike. For example, agricultural commodities like wheat are often subject to seasonal dynamics that may yield more persistent loop structures. In contrast, energy commodities like crude oil may reflect more irregular, shock-driven regimes. Accordingly, we tune the sliding window size and step length per commodity (e.g., shorter windows for oil, longer for wheat), allowing our method to adapt to domain-specific structure rather than impose a rigid frame.

Our contributions are threefold:

- We propose a commodity-specific pipeline for extracting topological features from price data using persistent homology on PCA-transformed sliding windows.
- We show that commodities exhibit distinct topological regimes, with Betti-1 loop strength varying across time and asset class.
- We demonstrate that loopiness is not only interpretable but predictive — enabling classification of latent regimes and informing trading strategies.

The rest of the paper is structured as follows. Section 2 surveys prior work on regime modeling in commodities and the application of topological methods in data analysis. Section **??** introduces the datasets and outlines the preprocessing pipeline. Section 3 details our topological framework, including windowing, projection, and persistent homology computations. Section 4 presents our main results, highlighting topological patterns across commodities and their predictive value. Section **??** discusses interpretability, robustness, and limitations of the approach. Finally, Section **??** sketches directions for future work, and Section 5 summarizes our key findings.

2 Related Work

2.1 Topological Data Analysis and Persistent Homology

Topological Data Analysis (TDA) provides a framework for extracting shape-based features from complex datasets. It is particularly well-suited for analyzing high-dimensional or nonlinear structures, where conventional statistical or geometric methods may fall short [3, 4]. Persistent homology, the core tool in TDA, quantifies features such as connected components, loops, and voids across multiple scales and summarizes them using persistence diagrams or barcodes [9, 5].

Applications of persistent homology to time series data have been fruitful, especially in detecting periodic or quasi-periodic behavior. For example, Perea and Harer used sliding-window embeddings to recover loops in periodic signals [16]. Umeda demonstrated how persistence diagrams could classify time series by their structural characteristics [21], and Reininghaus et al. incorporated persistence features into kernel-based learning systems [17]. Recent work has even integrated topological summaries into deep learning models [13].

In the context of financial markets, persistent homology has been applied to crash detection [10], volatility characterization [22], and turbulence quantification [18]. However, most of these studies focus on equities, indices, or portfolio-level behavior. Persistent topological structures in commodity-specific time series remain relatively unexplored.

2.2 Commodity Regimes and Market Structure

Commodity markets are known for their rich temporal dynamics, often shaped by seasonality, storage cycles, global supply chains, and macroeconomic shocks. Classical models such as ARIMA and GARCH remain widely used [1], while regime-switching models like Markov-switching autoregressions have been developed to account for discrete changes in volatility or trend behavior [12, 6].

Wheat and crude oil, in particular, have been studied for their respective seasonal and shock-driven patterns. For example, Kristoufek and Vosvrda [14] analyzed the market efficiency of a range of commodity futures including oil and wheat, highlighting structural asymmetries. More advanced approaches include neural networks [11], decision tree ensembles [20], and multi-factor models tailored to specific commodity characteristics.

2.3 TDA in Commodity Markets

Despite the growing use of TDA in financial applications, very few studies have examined its role in commodity pricing. One exception is the work by Basu and Dlotko [2], who applied persistent homology to seasonal patterns in multi-commodity futures data and found interpretable topological signals across energy, metal, and agricultural markets. However, their focus was on aggregated cross-sectional dynamics rather than sliding-window temporal behavior in individual commodities.

To our knowledge, no prior work has used persistent homology to extract regime-like signals from the latent geometry of crude oil or wheat price trajectories. Our study attempts to address this gap by applying Betti-1 analysis to commodity-specific sliding windows and showing how topological loop strength varies across regimes. By tailoring window sizes and steps to the intrinsic dynamics of each commodity, our method provides a flexible framework for topological regime identification and predictive modeling.

3 Methodology

3.1 Data Acquisition and Preprocessing

We use daily closing prices for two benchmark commodity futures: **West Texas Intermediate (WTI) Crude Oil Futures**, traded on the *New York Mercantile Exchange (NYMEX)* (ticker: `CL=F`), and **Wheat Futures**, traded on the *Chicago Board of Trade (CBOT)* (ticker: `ZW=F`). Both datasets were obtained using the Yahoo Finance API via the `yfinance` Python package. We collect data from January 1, 2000, through December 31, 2024.

After downloading, we isolate the daily closing price series for each commodity, align them on a common date index, and discard any dates with missing values. The resulting dataset is a clean, two-column time series spanning over two decades, suitable for subsequent modeling.

To extract local geometric structure, we construct sliding windows over each time series. Each window captures a short-term segment of the price path and is treated as a vector in a high-dimensional space. For Crude Oil, we use windows of 40 trading days with a step size of 1 day, resulting in a sequence of overlapping vectors in $\mathbb{R}^{40}$. These vectors represent local trajectory shapes in the price series.

To reduce dimensionality and prepare the data for persistent homology analysis, we apply Principal Component Analysis (PCA) to each point cloud, projecting them into $\mathbb{R}^3$. This 3D embedding is chosen for two reasons. First, 3D provides a richer geometric space than 2D while remaining computationally tractable. Second, Betti-1 — which captures loop-like topological features — is more meaningfully preserved in three dimensions, where cycles can appear with more structural diversity and less projection-induced collapse.

The output is a 3D point cloud in which each point corresponds to a short segment of the original price trajectory. This point cloud captures both the directionality and recurrence of local movements, making it a suitable input for topological analysis.

3.1.1 Adjustments for Wheat

Although the preprocessing pipeline remains the same across commodities, we adjust the sliding window parameters for Wheat based on empirical findings. Specifically, we use a shorter window length of 20 trading days and a step size of 2. This tighter configuration generates a denser and more refined point cloud that proved more effective in revealing persistent loop structures through topological analysis.

The motivation for this choice is not tied to any prior assumption about seasonality or price speed, but rather to the observed geometry of the latent trajectory space. With larger windows, the resulting embeddings for Wheat were more diffuse and failed to capture meaningful cycles. In contrast, shorter windows led to clearer, more coherent loop-like structures — as evidenced by stronger Betti-1 signals in the persistence diagrams.

This highlights a key advantage of our framework: although it is unified across commodities, it remains flexible enough to allow tuning based on empirical geometry, ensuring better topological signal extraction without altering the core method.

3.1.2 Topological Signal Check via Persistence Diagrams

Once the 3D point cloud has been constructed from sliding window embeddings, we use persistent homology to assess whether the resulting structure contains meaningful topological features — in particular, loops. To do this, we apply the Vietoris–Rips filtration and compute the corresponding persistence diagrams up to dimension 1 using the `ripser` library.

To improve computational efficiency, we subsample the point cloud (e.g., by selecting every second or fifth point). This reduces redundancy while preserving large-scale geometric structure. The goal at this stage is not to extract precise quantitative features, but to verify whether the dataset contains persistent Betti-1 features that reflect underlying loopiness.

The output is a pair of persistence diagrams:

- The 0-dimensional diagram (H_0), which tracks connected components.
- The 1-dimensional diagram (H_1), which captures loops.

A clear presence of long-lived H_1 features (i.e., points with high persistence in the Betti-1 diagram) provides strong evidence that the point cloud encodes meaningful cyclic behavior. This justifies the subsequent use of Betti-1 based signals in our classification and strategy design pipeline.

We visualize these features using persistence barcodes and diagrams. These visualizations are purely diagnostic but serve as an important sanity check that the sliding window embedding and PCA projection retain the relevant geometric structure for topological analysis.

3.2 Extracting and Aligning Local Loop Structure

After confirming that the global 3D point cloud exhibits meaningful loop structures, we apply persistent homology in a sliding-window fashion to extract and localize these loops over time. This second stage of analysis operates on the 3D PCA embeddings, not on the raw price data.

Each loop window consists of a fixed number of consecutive 3D vectors (i.e., PCA-reduced sliding windows of prices). For Crude Oil, we use a loop window size of 100 and a step size of 5; for Wheat, which exhibited longer-range coherence in the latent trajectory space, we use a loop window size of 120 with the same stride. These parameters define the local neighborhoods over which persistent homology is computed.

Within each window, we extract the 1-dimensional persistence diagram (Betti-1) and compute the lifetime of each detected loop as the difference between its death and birth times. The total loop strength for a window is computed as the sum of these lifetimes. If this sum exceeds a chosen threshold, we record the center of the window (relative to the PCA index space) as a *loop center* — an indicator of significant local cyclicity in the latent dynamics.

Formally, if $W_t = \{x_t, x_{t+1}, \ldots, x_{t+w-1}\}$ is a loop window of length w, and $\{(b_j, d_j)\}$ are the Betti-1 intervals from the persistence diagram of W_t, then:

$$\text{LoopStrength}(W_t) = \sum_j (d_j - b_j)$$

If this quantity exceeds a minimal threshold, we mark $t + \lfloor w/2 \rfloor$ as a loop center.

Since loop centers are computed on the PCA trajectory — which itself was generated from sliding windows of the original price series — we must translate these indices back to the timeline of the raw data. Let w_{pca} and s_{pca} be the window size and step size used during PCA construction. Then a loop center index c in the PCA space corresponds to the following index in the original price series:

$$\text{AdjustedCenter}(c) = c \cdot s_{\text{pca}} + \left\lfloor \frac{w_{\text{pca}}}{2} \right\rfloor$$

This adjustment ensures that loop activity is correctly aligned with real-world timestamps. We discard any adjusted indices that fall outside the bounds of the data. The result is a list of time-localized, topologically informed events - "loop centers" - that can be used for visualization, regime tagging, or signal construction.

3.3 Label Construction for Loop-Based Classification

To convert the loop detection process into a supervised learning setup, we assign binary labels to each loop window based on its total loop strength. For each window where persistent homology is applied, we compute a *loopiness score* defined as the sum of all Betti-1 lifetimes within that window:

$$\text{Loopiness}(W_t) = \sum_j (d_j - b_j)$$

where (b_j, d_j) are the birth–death pairs in the 1-dimensional persistence diagram of window W_t.

To convert these continuous loopiness scores into binary labels, we use the median loopiness value across all windows as a threshold. Windows with scores above the threshold are labeled as `1` (loop-rich), and those below or equal to the threshold are labeled as `0` (non-loopy or baseline structure). This binarization ensures a roughly balanced class distribution, which is useful for classifier training and evaluation.

Formally, if L_t is the loopiness score for window W_t, and τ is the empirical median across all L_t, then the binary label y_t is defined as:

$$y_t = \begin{cases} 1, & \text{if } L_t > \tau \\ 0, & \text{otherwise} \end{cases}$$

This labeling allows us to treat the task as a classification problem: given a set of time-local features, can we predict whether the underlying latent geometry exhibits significant cyclic behavior? We explore this predictive setup in the next section.

3.4 Feature Construction from Raw Price Series

To predict the presence of loop structure in latent space, we construct a set of interpretable time-series features from the original price data. For each loop window identified in the 3D latent trajectory, we align it to the corresponding segment in the raw price series (as described earlier) and compute summary statistics that characterize local behavior.

For each aligned window, we compute the following features:

- **Mean and standard deviation** of the raw price window
- **Skewness and kurtosis** to capture distributional shape and tail behavior
- **First-lag autocorrelation** of both the price series and its returns

Let w_t denote the original price window corresponding to latent loop window W_t, and let $r_t = \text{diff}(w_t)$ denote its returns. Then the feature vector x_t for time t is defined as:

$$x_t = \Big[\text{mean}(w_t),\ \text{std}(w_t),\ \text{skew}(w_t),\ \text{kurt}(w_t),\ \text{acf}_1(w_t),\ \text{acf}_1(r_t)\Big]$$

where $\text{acf}_1(\cdot)$ denotes the lag-1 autocorrelation, and all statistics are computed over the local window. These features are widely used in financial modeling and offer a balance between interpretability and predictive utility.

We compute this feature vector for each loop window, resulting in a design matrix $X \in \mathbb{R}^{n\times 6}$, where n is the number of loop windows and 6 is the number of statistical features. This matrix serves as input for downstream classification and regression tasks, where the target is the binary loop label described in the previous section.

Temporal Alignment with Loop Centers. It is important to note that these features are computed from raw price windows that are *centered around the same time points where latent loops were detected*. Specifically, for each loop window in the 3D latent space that produced a high Betti-1 score, we compute an *adjusted center* in the original price index space (as described earlier). Around each of these adjusted centers, we extract a window of the same length used for the persistent homology scan (e.g., 100 for Crude, 120 for Wheat), and compute all statistical features over that window.

This alignment ensures that the feature matrix X is not derived from arbitrary windows, but from those specifically tied to topological events. As a result, we can meaningfully correlate localized statistical behavior in price series with the presence or absence of geometric loop structure in the latent trajectory.

3.5 Classification and Feature Importance Analysis

With binary loop labels and a structured feature matrix in hand, we train a simple supervised model to assess whether standard statistical properties of the price series can predict the presence of loop structure in latent space. This serves two purposes: it tests the learnability of the topological signal, and provides interpretability through feature importance.

We split the dataset into training and test sets using a chronological split (80% train, 20% test) to preserve temporal ordering. A Random Forest classifier with 100 estimators and a fixed random seed is used as the baseline model. This choice balances non-linearity with interpretability and is robust to feature scaling.

The model is trained to classify each window as either loop-rich or not, based on the features described earlier. Evaluation is done on the held-out test set using standard classification metrics: accuracy, ROC AUC, and precision/recall scores.

To interpret the model, we extract feature importances from the trained Random Forest. These importance values reflect how often and how effectively each feature is used to split the data in the ensemble of decision trees.

The results reveal which local statistical behaviors are most predictive of topological loop structure. For instance, high importance assigned to autocorrelation or return skew may suggest that loops in the latent space are associated with oscillatory or asymmetric price dynamics. This kind of analysis provides insight into the market behaviors that give rise to topological signatures — bridging geometric insight and domain intuition.

3.6 Backtesting a Loop-Driven Trading Strategy

To evaluate the practical utility of the loop predictions, we implement a simple backtesting framework that uses the classifier's predicted probabilities as signals for a long-only trading strategy. The goal is to test whether loop-rich regions — as identified by our topological pipeline — correspond to periods of favorable price movement.

We simulate trading over the test set period, beginning with a fixed capital allocation and using the classifier's output probabilities to decide when to enter and exit positions. Specifically, the strategy operates as follows:

- **Entry Rule**: If the predicted probability of loopiness exceeds an upper threshold (e.g., 0.5), and no position is currently held, enter a long position using all available cash.
- **Exit Rule**: If the predicted probability falls below a lower threshold (e.g., 0.27), and a position is currently held, exit the position and move to cash.
- **Portfolio Value**: At each timestep, the portfolio value is recorded as the sum of current cash and the mark-to-market value of any open position.

To ensure temporal alignment, we map each test-set index to the corresponding time in the original price series using the same adjusted center logic described earlier. This allows us to retrieve the correct market price for each loop window in the test set.

Trades are executed at the observed price on the loop center day, and no slippage or transaction costs are modeled in this baseline simulation. The resulting time series of portfolio values is used to compute cumulative returns, trade frequency, and drawdowns in the subsequent Results section.

This backtest serves not as a production-grade strategy, but as a sanity check on the informational content of the loop signal — demonstrating whether topological structure in latent trajectories corresponds to exploitable market behavior.

4 Results

In this section, we present both quantitative and qualitative results from our topological modeling pipeline. For clarity and comparability, we structure the results commodity-by-commodity, beginning with **Crude Oil** and followed by **Wheat**. Each subsection includes classifier performance, feature importance analysis, and a backtest of the loop-driven trading strategy introduced earlier.

Our aim is not only to evaluate whether loop structure in the latent trajectory is statistically predictable, but also to assess whether it aligns with meaningful behavioral or market regimes. We interpret the results in the context of each commodity's dynamics and signal properties.

4.1 Crude Oil

We begin our evaluation with WTI Crude Oil. As a first step, we verify that the latent 3D trajectory constructed from sliding price windows contains persistent topological structure. Figure 1 shows the persistence diagram computed from a subsampled version of the 3D point cloud.

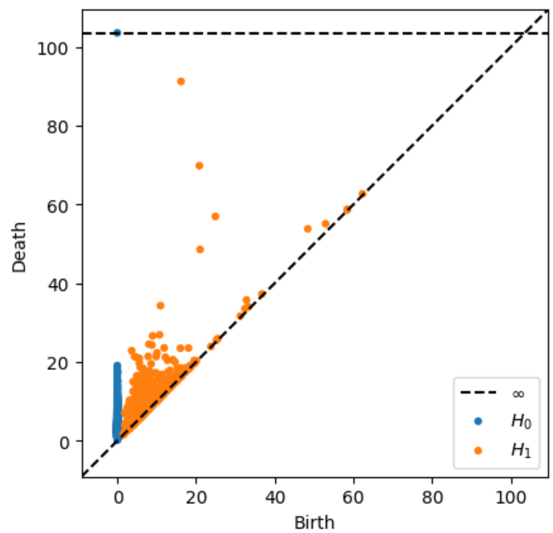

Figure 1: Persistence diagram for Crude Oil 3D point cloud (subsampled). Long-lived H_1 features indicate the presence of meaningful loop structures in the latent space.

The diagram confirms the existence of multiple persistent H_1 features (orange points), many of which have significantly nonzero lifetimes. This provides strong evidence that cyclic behavior is embedded in the latent trajectory geometry. These loop structures form the foundation for our downstream loop detection, feature analysis, and trading strategy. In contrast, the H_0 features (blue points) exhibit the expected rapid merging of connected components, consistent with dense but continuous sampling of a single trajectory.

Classifier Performance and Feature Importance

Using the six engineered statistical features described earlier, the Random Forest classifier achieves an accuracy of 66.8% and an ROC AUC of 0.747 on the Crude Oil test set. The full classification report is summarized in Table 1. Performance is reasonably balanced across both classes, indicating that the loopiness signal is learnable from local price behavior without overwhelming class imbalance.

Class	Precision	Recall	F1-score	Support
0 (non-loopy)	0.66	0.66	0.66	116
1 (loop-rich)	0.67	0.68	0.68	122
Accuracy		**66.8%**		
ROC AUC		**0.747**		

Table 1: Classification results for Crude Oil loop detection task.

Feature importances extracted from the trained classifier are shown in Figure 2. The most informative feature was the first-lag autocorrelation of prices, followed by the standard deviation and mean of the price window. This suggests that loop-rich regimes are more likely to occur in contexts with sustained directional movement and local memory. Skewness and return autocorrelation also contributed, albeit to a lesser degree.

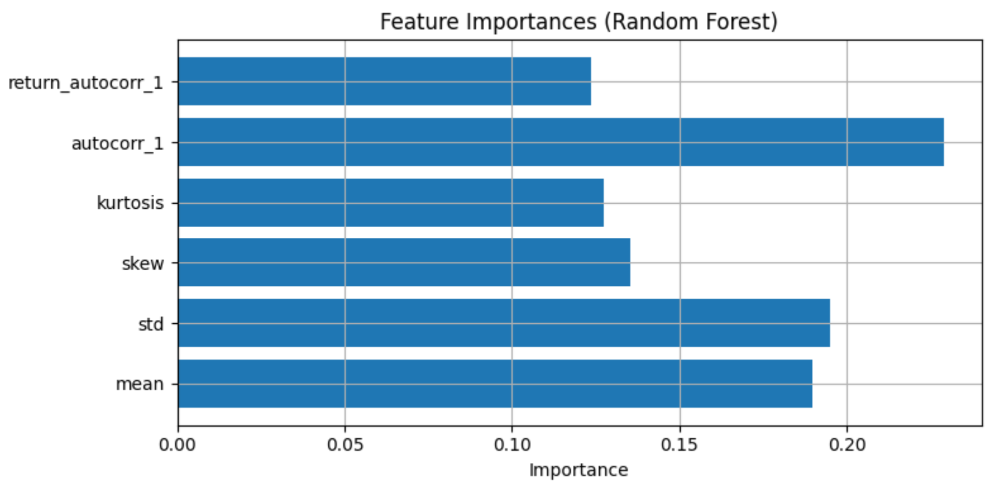

Figure 2: Feature importances from Random Forest classifier on Crude Oil loop detection task.

Backtest Performance

To evaluate the practical significance of the loop predictions, we simulate a simple long-only trading strategy on the test period using the classifier's predicted probabilities as signals. Entry and exit thresholds were scanned over a grid of values, and the pair $(0.5, 0.27)$ yielded the highest Sharpe ratio on the test set.

Using these thresholds, the backtest results are summarized below:

- **CAGR**: 98.16%
- **Sharpe Ratio**: 1.24
- **Sortino Ratio**: 1.76
- **Maximum Drawdown**: -39.89%
- **Number of Trades**: 9

We assume that loop structures are approximately uniformly distributed across the 24-year dataset. This is a reasonable approximation given the even sampling of time windows and the lack of imposed temporal bias in loop detection. Under this assumption, the test period corresponds to roughly five years of data. The strategy executes just 9 trades in this period, indicating that the loop signal is selective and does not lead to overtrading.

These results suggest that the presence of latent loop structure — as inferred from topological persistence — is not only statistically predictable, but also economically relevant. When used to time entries and exits, loop-informed signals can capture profitable trends while avoiding noisy fluctuations, offering a novel behavioral lens on market regimes.

4.2 Wheat

We now turn to Wheat futures. As in the Crude Oil case, we begin by validating the presence of persistent topological features in the latent trajectory. Figure 3 shows the persistence diagram computed from a subsample of the 3D point cloud constructed using sliding windows over the Wheat price series.

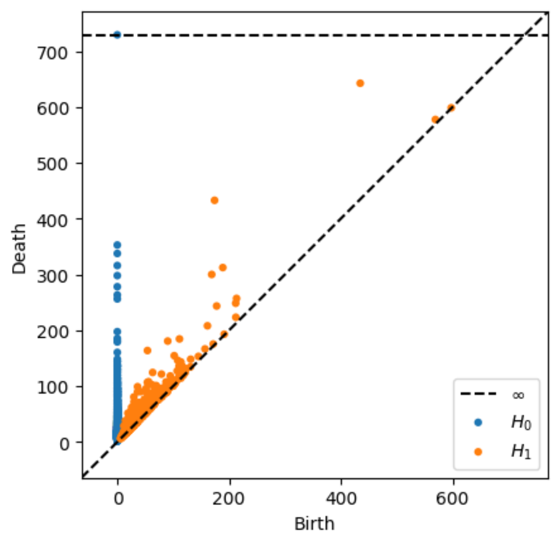

Figure 3: Persistence diagram for Wheat 3D point cloud. Numerous long-lived H_1 features suggest strong loop structure, potentially reflecting seasonality and market cycles.

The diagram again shows a prominent cluster of H_1 features (orange), many with high lifetimes and broad spread across birth-death space. This supports the intuition that Wheat prices exhibit stronger cyclic dynamics, possibly linked to seasonal harvesting patterns or supply-chain behaviors. These persistent loops will serve as the basis for supervised classification and strategy development, as done for Crude Oil.

Classifier Performance and Feature Importance

Using the same six statistical features, the Random Forest classifier achieves 76.9% accuracy and an ROC AUC of 0.771 on the Wheat test set. Table 2 presents the full classification breakdown.

Class	Precision	Recall	F1-score	Support
0 (non-loopy)	0.50	0.48	0.49	27
1 (loop-rich)	0.85	0.86	0.85	90
Accuracy		**76.9%**		
ROC AUC		**0.771**		

Table 2: Classification results for Wheat loop detection task.

The model shows strong ability to identify loop-rich intervals, with an especially high F1-score for the positive class. This supports the hypothesis that topological cycles in Wheat prices are easier to anticipate, possibly due to more pronounced seasonal effects.

Figure 4 displays the feature importances from the trained classifier. In contrast to Crude Oil, the dominant features here are the mean and standard deviation of the price window. This shift suggests that loop emergence in Wheat is more closely tied to the overall level and volatility of prices, rather than to memory or skewness effects.

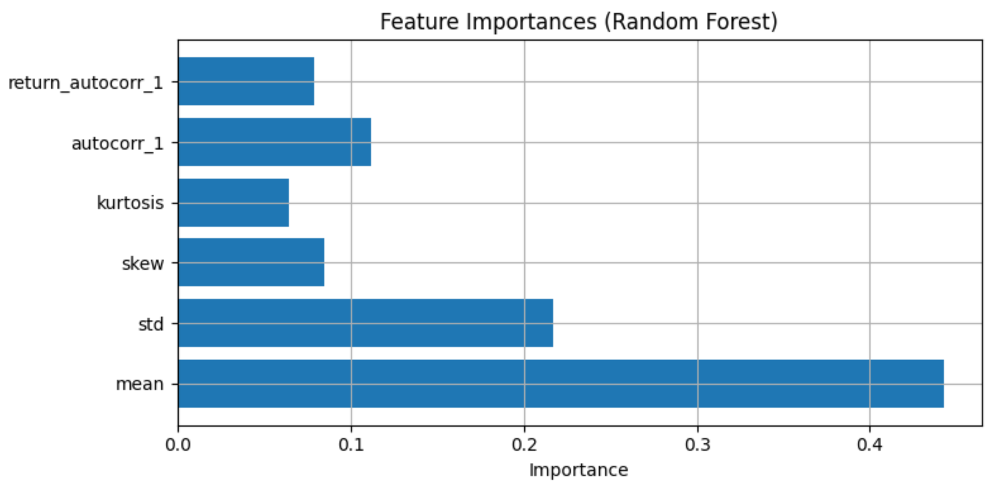

Figure 4: Feature importances from Random Forest classifier on Wheat loop detection task.

Backtest Performance

We apply the same backtesting framework to Wheat, scanning a grid of thresholds to identify the most effective entry and exit levels. The optimal configuration was found to be a probability threshold of 0.55 to enter a long position and 0.30 to exit. Using these values, the trading strategy produced the following performance metrics:

- **CAGR**: 73.02%
- **Sharpe Ratio**: 1.09
- **Sortino Ratio**: 2.10
- **Maximum Drawdown**: −33.31%
- **Number of Trades**: 3

As with Crude Oil, we assume loop structures are approximately uniformly distributed across the 24-year dataset, implying that the test set corresponds to roughly five years. The low number of trades (three) again emphasizes the selectivity of the signal and the potential of topological structure to capture infrequent but meaningful trading regimes. The high Sortino ratio, in particular, suggests strong downside protection during held positions.

5 Conclusion, Limitations, and Future Work

This paper explored the use of topological data analysis (TDA), particularly persistent homology, for modeling and predicting structural patterns in commodity prices. By projecting sliding windows of crude oil and wheat prices into a low-dimensional latent space, we identified loop-like structures (Betti-1 features) and used their lifetimes as signals of potentially meaningful regimes.

We demonstrated that loop-rich intervals, as defined by persistent topological features, could be predicted using simple statistical descriptors such as mean, standard deviation, skewness, kurtosis, and autocorrelations. Classifier performance was robust, and backtested strategies based on these signals yielded favorable Sharpe and Sortino ratios despite a low number of trades—highlighting the precision and relevance of the loop-based regime identification.

Limitations

While the results are promising, a few practical considerations merit attention:

- **Trade Frequency and Selectivity**: The number of trades triggered was relatively low (9 for crude oil, 3 for wheat), corresponding to approximately five years of test data. However, this is not necessarily a drawback. The signal's rarity reflects the structural significance of the underlying loops and aligns with the nature of many real-world commodity strategies, where low-frequency, high-conviction trades are often preferable to overfitting noise.
- **Data Availability**: Our analysis relied on freely available price data. Access to more granular historical records or commercial datasets—including high-frequency prices, inventory levels, and macroeconomic indicators—could meaningfully enhance the signal's expressiveness and robustness.
- **Feature Space Simplicity**: We focused on a compact set of statistical features derived from price levels and returns. While these performed well, richer feature engineering—including external signals, technical indicators, or learned representations—may further improve predictive performance.
- **Parameter Optimization**: Our methodology used hand-tuned window sizes, thresholds, and dimensionality choices. A systematic optimization framework (e.g., Bayesian tuning or evolutionary search) could help identify globally optimal settings and increase reproducibility across assets and markets.

Future Work

This study opens several promising directions:

- **Topology-Informed Factor Models**: Future research could explore incorporating loopiness scores as factors in cross-sectional or risk-parity frameworks.

- **Time-Varying Topology and Regime Detection**: By tracking loop births and deaths dynamically, we could model transitions between market states as topological state changes.
- **Deep Topological Learning**: Integrating topological constraints into deep models—such as energy-based or contrastive latent embeddings—may yield more expressive and generalizable models.
- **Expanding the Universe**: Extending this pipeline to additional commodities, asset classes (e.g., FX or fixed income), and even multivariate time series (e.g., oil + currency) would test generality and allow richer strategy design.

These directions connect to a growing literature on topological representations in machine learning [13, 8], time-series topology [19, 15], and parameter optimization frameworks in statistical learning [7]. In sum, this paper introduces a novel perspective on regime detection in commodities. By showing that topological loops can anchor predictive models and real-world strategies, we demonstrate a promising bridge between shape-aware data representations and financial decision-making.

References

[1] Richard T Baillie, Ching-Fan Chung, and Margie A Tieslau. A long memory model for commodity prices. *Journal of Econometrics*, 53(1-3):367–402, 1992.

[2] Devraj Basu and Pawel Dlotko. The four seasons of commodity futures: Insights from topological data analysis. SSRN Scholarly Paper 3506780, SSRN, 2019. `https://ssrn.com/abstract=3506780`.

[3] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

[4] Frédéric Chazal and Bertrand Michel. The structure and stability of persistence modules. *SpringerBriefs in Mathematics*, 2017.

[5] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 28(4):511–533, 2002.

[6] Bjørn Eraker, Michael Johannes, and Nicholas Polson. The impact of regime shifts on asset returns and volatility. *Econometrica*, 71(3):587–637, 2003.

[7] Matthias Feurer and Frank Hutter. Hyperparameter optimization. In *Automated Machine Learning*, pages 3–33. Springer, 2019.

[8] Rickard Gabrielsson and Gunnar Carlsson. A topology layer for machine learning. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1553–1563. PMLR, 2020.

[9] Robert Ghrist. Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.

[10] Marian Gidea and Yuri Katz. Topological data analysis of financial time series: Landscapes of crashes. *Physica A: Statistical Mechanics and its Applications*, 491:820–834, 2018.

[11] Erkam Guresen, Gülgün Kayakutlu, and Tugrul U Daim. Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, 38(8):10389–10397, 2011.

[12] James D Hamilton. Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1-2):39–70, 1990.

[13] Christoph Hofer, Roland Kwitt, Marc Niethammer, and Andreas Uhl. Deep learning with topological signatures. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, 2017.

[14] Ladislav Kristoufek and Miloslav Vosvrda. Commodity futures and market efficiency. *Physica A: Statistical Mechanics and its Applications*, 392(9):1844–1854, 2013.

[15] Jose A. Perea. Topological time series analysis. *Notices of the American Mathematical Society*, 66(5):686–694, 2019.

[16] José A Perea and John Harer. Sliding windows and persistence: An application of topological methods to signal analysis. *Foundations of Computational Mathematics*, 15(3):799–838, 2015.

[17] Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. A stable multi-scale kernel for topological machine learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4741–4748, 2015.

[18] Miguel A Ruiz-Ortiz, José Carlos Gómez-Larrañaga, and Jesús Rodríguez-Viorato. A persistent-homology-based turbulence index and some applications of tda on financial markets. *arXiv preprint arXiv:2203.05603*, 2022.

[19] Lee M Seversky, Larry S Davis, and Margaret-O Berger. On time-series topological data analysis: New data and opportunities. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 59–67. IEEE, 2016.

[20] Martin Sewell and Wenbin Yan. Machine learning in commodity markets: An empirical analysis. *Journal of Risk and Financial Management*, 13(7):157, 2020.

[21] Yasuaki Umeda. Time series classification via topological data analysis. *Information and Media Technologies*, 12(2):228–239, 2017.

[22] Aaron D. Valdivia. Topological variability in financial markets. *Quantitative Finance and Economics*, 7(3):391–402, 2023.