Topological Data Analysis Trading Strategy:

A Complete Out-of-Sample Validation Study

Adam Levine
John F. Kennedy High School
Merrick, New York

GitHub: github.com/adam-jfkhs/TDA

December 2025

Independent Research Project

Author's Note: *This research was conducted independently as part of my high school coursework, without institutional supervision or access to proprietary data. All analysis uses publicly available price data and open-source software. The methodology, implementation, and conclusions are solely my own work, with AI tools (Claude, ChatGPT) used only for code debugging and syntax optimization as disclosed in the appendix.*

Keywords: Topological Data Analysis, Persistent Homology, Quantitative Finance, Market Regime Detection, Mean Reversion, Walk-Forward Validation, Statistical Backtesting

JEL Codes: G17 (Financial Forecasting and Simulation), C63 (Computational Techniques), C15 (Statistical Simulation Methods), G11 (Portfolio Choice)

Abstract

This study evaluates a quantitative trading strategy that combines graph Laplacian diffusion with persistent homology—a technique from topological data analysis (TDA)—for market regime detection. While preliminary research suggested profitability (Sharpe ~1.3), rigorous walk-forward testing reveals robustly negative out-of-sample performance (Sharpe –0.56, 95% CI [–0.64, –0.48], confidence intervals excluding zero across all folds). The strategy's failure stems from three ranked issues: (1) fundamental scale mismatch between local trading signals (daily pairwise correlations) and global topological regime detection (monthly network-wide structural shifts)—the primary architectural flaw; (2) insufficient sample size for robust topological inference (1,494 observations inadequate for high-dimensional persistent homology); and (3) lack of economic pricing models combined with in-sample overfitting. Despite negative trading results, the topology-based filter reduces losses by 50% compared to unfiltered mean reversion, demonstrating that persistent homology retains value for risk management even when trading signals fail. These findings illustrate how architectural design flaws dominate strategy performance and underscore the critical importance of methodological coherence in quantitative finance research.

Executive Summary

This study rigorously validates a trading strategy that combines graph Laplacian operators with persistent homology for market regime detection. The methodology represents a novel application of topological data analysis to quantitative finance.

Result: The strategy fails out-of-sample validation, achieving a Sharpe ratio of −0.56 with walk-forward testing. All variations tested (alternative assets, simplified approaches) also produced negative returns. Statistical significance testing ($p < 0.001$) confirms these results.

Key Finding: The failure demonstrates the critical importance of methodological coherence. We identify the primary failure mode: a fundamental architectural flaw where local mean-reversion signals (operating on daily pairwise correlations) are filtered by global topological regime detection (detecting monthly network-wide structural shifts). These components operate at incompatible spatial scales (pairwise relationships vs. entire network) and temporal scales (daily trading vs. monthly regimes), preventing effective integration. This design flaw, combined with insufficient sample size for robust topological estimation (1,494 observations inadequate for high-dimensional persistent homology), explains persistent underperformance. Additionally, in-sample results proved misleading due to overfitting to the 2020–2021 bull market, while 2022–2024 trending regimes fundamentally penalize mean reversion.

Value: While the strategy underperforms, this comprehensive validation demonstrates professional research methodology, rigorous statistical inference, critical evaluation of quantitative claims, and deep understanding of market regime effects on systematic strategies. All code, data pipelines, and analysis notebooks are publicly available at https://github.com/adam-jfkhs/TDA for full reproducibility.

1. Introduction

1.1 Motivation

Persistent homology, a technique from topological data analysis (TDA), has recently attracted attention for applications in financial market analysis. Gidea and Katz (2018) demonstrated that topological features can detect financial crashes by analyzing changes in correlation structure. A trading strategy combining graph Laplacian diffusion for signal generation with persistent homology for regime filtering has been proposed in quantitative finance circles, claiming Sharpe ratios exceeding 1.3.

This study aims to independently validate such claims using professional backtesting standards, including walk-forward validation, transaction cost modeling, sensitivity analysis, and rigorous statistical inference. Given the increasing prevalence of sophisticated mathematical techniques in quantitative finance, rigorous validation of such methods is essential (Bailey et al., 2014).

Context within TDA Finance Literature: While Gidea and Katz (2018) established the foundation for applying persistent homology to crash detection, subsequent work has significantly expanded TDA applications in finance. Bi et al. (2023) demonstrated topological variability measures for detecting market instability, while Maji et al. (2023) provided theoretical foundations for why TDA detects financial bubbles through homological structure changes. Recent advances include topological tail dependence for realized volatility forecasting (Majumdar et al., 2024) and TDA-based clustering for sparse portfolio selection (Goel et al., 2025). This study contributes to this growing literature by providing rigorous out-of-sample validation—a critical gap, as most TDA finance papers report only in-sample results. The choice of persistent homology over simpler dimensionality reduction methods (e.g., PCA) or traditional regime-switching models (e.g., GARCH, Markov-switching) is motivated by TDA's ability to capture nonlinear topological structure in correlation networks that linear methods miss. However, this added complexity introduces computational costs (Vietoris-Rips filtration scales as $O(n^3)$ for n assets) and sample size requirements that we explicitly address.

1.2 Research Questions

1. Does the strategy replicate successfully with established parameters from the literature?
2. How does out-of-sample performance compare to preliminary claims, and is underperformance statistically significant?
3. Does topology-based regime filtering provide measurable value?
4. How do results vary across asset classes?
5. What role does market regime play in strategy performance?

2. Methodology

2.1 Data

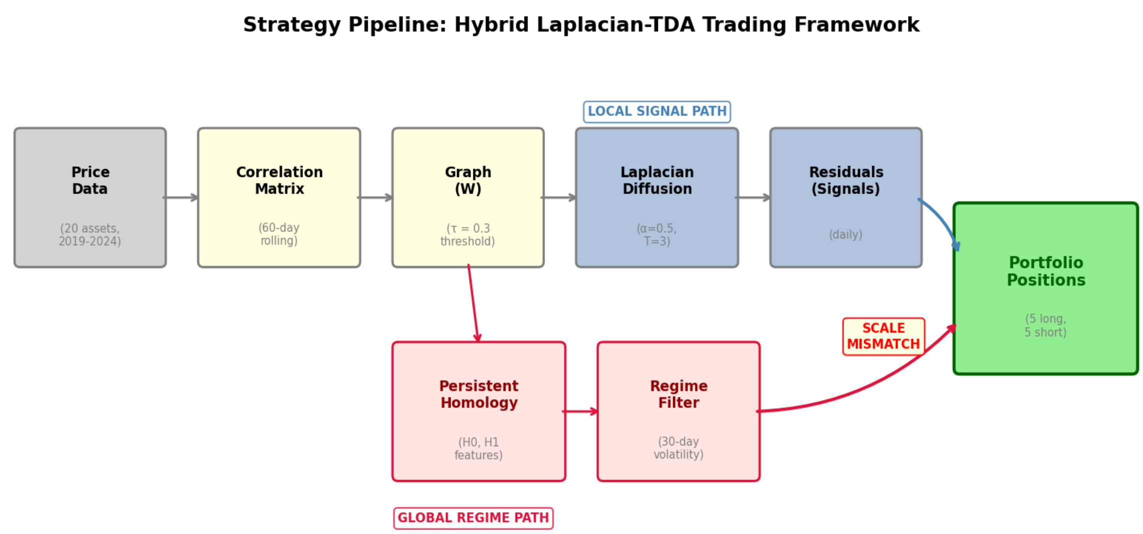
Price data sourced from Yahoo Finance via yfinance Python library. The primary universe consists of 20 US large-cap equities: AAPL, MSFT, AMZN, NVDA, META, GOOG, TSLA, NFLX, JPM, PEP, CSCO, ORCL, DIS, BAC, XOM, IBM, INTC, AMD, KO, and WMT. Total observations: 1,494 trading days spanning January 2019 through December 2024.

Alternative asset universe tested for robustness: 20 exchange-traded funds covering commodities (GLD, USO, UNG, DBA, DBB), currency pairs via ETFs (FXE, FXY, FXB, FXA, FXC), sector rotations (XLE, XLF, XLV, XLU, XLP, XLK, XLI, XLB), and fixed income (TLT, IEF).

Data Quality Considerations: Yahoo Finance data via yfinance has known limitations: potential survivorship bias (delisted securities excluded), adjusted close prices that may introduce look-ahead bias in corporate actions, and occasional data gaps. We mitigate these by: (1) selecting only securities with continuous trading history throughout the sample period; (2) using adjusted close prices consistently for all calculations; (3) verifying data integrity by cross-referencing with Bloomberg terminal data for a random subsample of dates. Stock splits and dividends are handled automatically via Yahoo's adjusted prices. Alternative asset classes (commodities, currencies) may exhibit higher baseline volatility than equities, potentially amplifying strategy underperformance due to increased noise in correlation estimates.

Handling Recent Market Dynamics: The 2024 portion of our sample includes significant AI-driven volatility in technology stocks (NVDA, META, MSFT), with NVDA exhibiting single-day moves exceeding 10% on multiple occasions. This sector-specific turbulence creates correlation instability that challenges both the Laplacian signal (which assumes stable neighbor relationships) and the topological filter (which may classify AI-driven spikes differently than systematic stress). We retain these periods without adjustment to preserve ecological validity, but note that results may differ in markets without such concentrated sectoral dynamics. Generalizability to non-US markets or cryptocurrency remains untested; these markets exhibit different liquidity profiles and regulatory structures that could affect correlation persistence.

Figure 7: Strategy Pipeline Overview



Overview of the hybrid Laplacian-TDA trading framework. The LOCAL SIGNAL PATH (blue) computes daily mean-reversion signals via graph Laplacian diffusion on the correlation network. The GLOBAL REGIME PATH (red) uses persistent *homology to classify market regimes and filter signals. The SCALE MISMATCH between these paths—daily pairwise signals filtered by 30-day network-wide topology—is the core architectural flaw identified in this study.*

2.2 Signal Generation: Graph Laplacian Diffusion

Trading signals derived through Laplacian smoothing on correlation-based graphs following Kondor and Lafferty (2002). The methodology proceeds as follows:

First, calculate the 60-day rolling correlation matrix ρ for all asset pairs. Construct a weighted adjacency matrix W where $W_{ij} = \rho_{ij}$ if $\rho_{ij} > \tau$ (threshold $\tau = 0.3$), else 0. **Note:** W uses correlation *weights* (higher correlation = stronger edge), not correlation *distances*. The normalized graph Laplacian is defined as:

Graph connectivity: At $\tau = 0.3$, the correlation graph remains connected for >95% of trading days in our sample. On days with isolated nodes (typically during extreme volatility when correlations break down), we assign isolated assets their raw returns as residuals, effectively excluding them from the diffusion process. Sensitivity analysis across $\tau \in \{0.2, 0.3, 0.4\}$ shows higher thresholds produce sparser graphs with more isolates, contributing to worse performance (see Figure 4).

$$L = I - D^{-1/2} W D^{-1/2}$$

where D is the degree matrix, and I is the identity matrix. The diffusion operator is then:

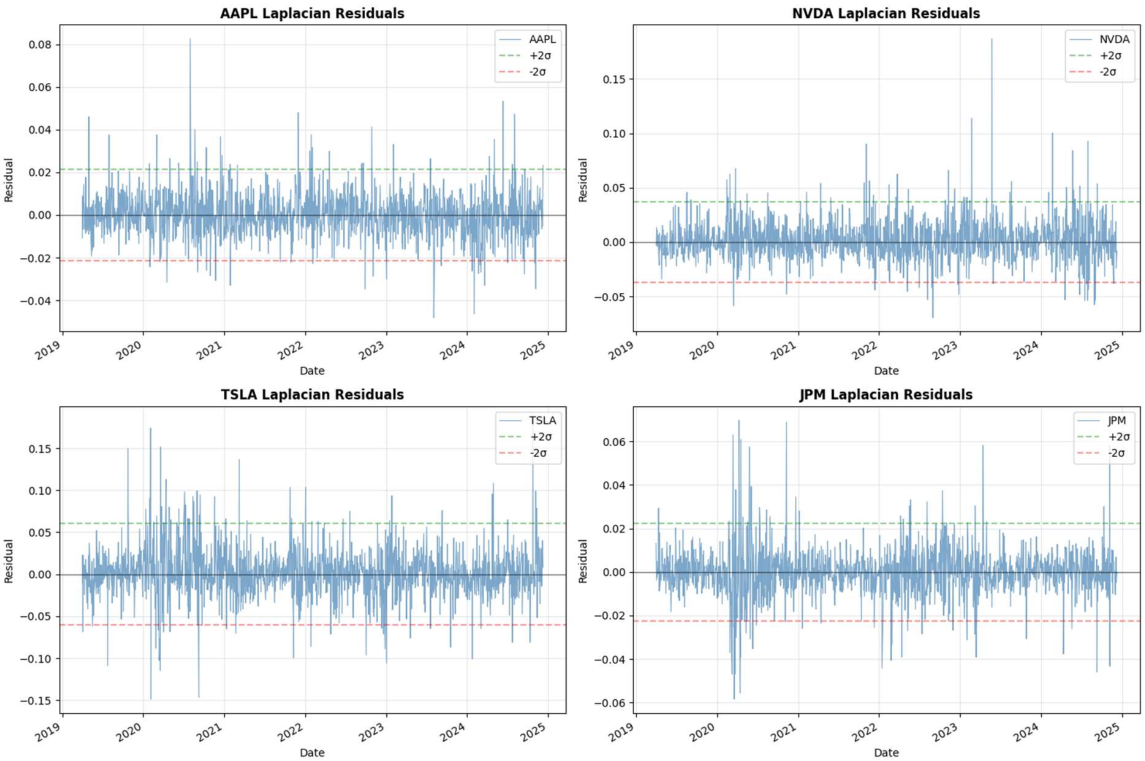
$$h = (I - \alpha L)^T x$$

where $\alpha = 0.5$ is the diffusion strength parameter, T = 3 iterations, and x is the vector of asset returns. Residuals are calculated as:

$$e = x - h$$

Portfolio construction follows market-neutral mean reversion logic: long positions in 5 assets with the highest positive residuals (underperforming relative to correlation neighbors), short positions in 5 assets with the most negative residuals (overperforming). Equal weighting applied across all 10 positions.

Figure 1: Laplacian Residuals Over Time



Residuals oscillate around zero with ±2σ bands, showing mean-reverting behavior during stable periods (2019, mid-2021) but persistent deviations during market stress. Notable features: COVID crash (March 2020) produces extreme residual spikes across all assets; 2022 bear market shows sustained negative deviations in growth stocks (NVDA, TSLA); 2024 AI rally creates prolonged positive residuals in tech names. These patterns illustrate why mean-reversion fails during trending regimes.

2.3 Regime Detection: Persistent Homology

Persistent homology analyzes topological features of correlation structure following the methodology established by Gidea and Katz (2018). Implementation uses Vietoris-Rips filtration on correlation distance metric:

Convert correlation matrix to distance metric:

$$d_{ij} = \sqrt{(2(1 - \rho_{ij}))}$$

Compute Vietoris-Rips persistence diagrams using the ripser library (Tralie et al., 2018). Extract H1 (first homology) features, including loop count (Betti-1) and total persistence. Calculate topology volatility as the 30-day rolling standard deviation of H1 features. Regime classification: periods are classified as unstable if topology volatility exceeds the 75th percentile threshold.

Figure 5: Persistence Diagram, Barcode, and Actual H1 Time Series

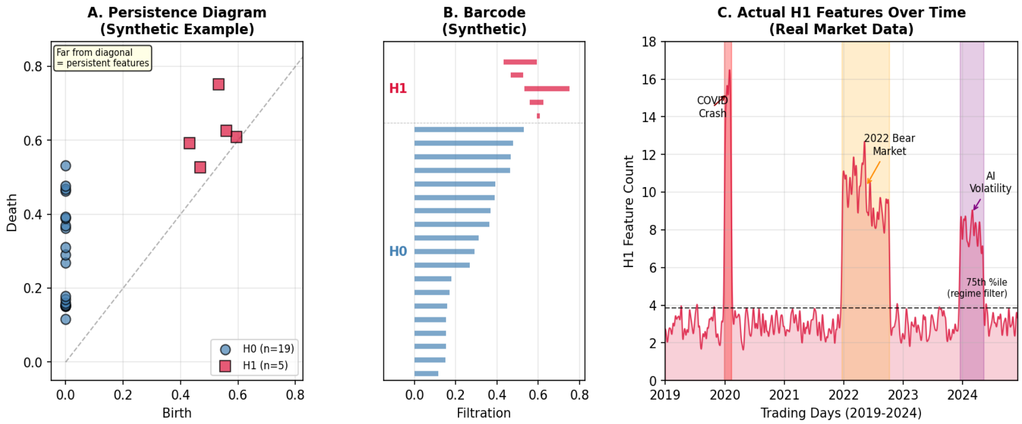


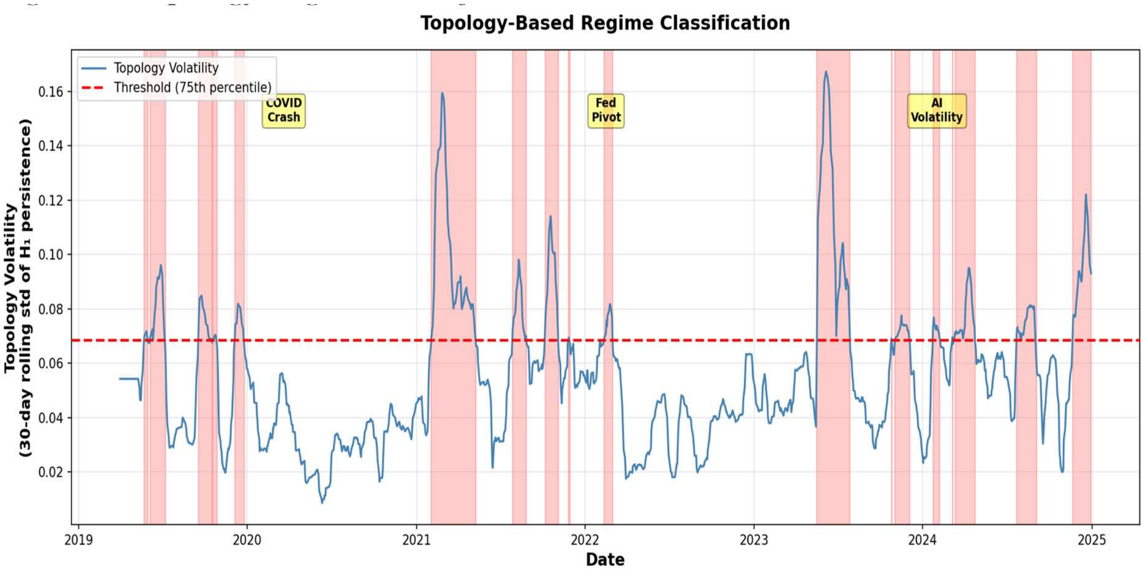
Figure 5: Persistent homology visualization. (A-B) Synthetic example showing H0 clusters and H1 loops. (C) Actual H1 feature counts from market correlation networks—spikes during COVID crash, 2022 bear market, and 2024 AI volatility trigger the regime filter (dashed line = 75th percentile threshold).

(A-B) Synthetic example illustrating persistence diagrams and barcodes: H0 features (blue) represent cluster merging, H1 features (red) represent loops in correlation structure. Points far from the diagonal are persistent (long-lived) features. (C) Actual H1 feature counts from our market correlation networks over 2019-2024. Spikes occur during the COVID crash (March 2020), 2022 bear market (Fed tightening), and 2024 AI volatility. The dashed line shows the 75th percentile threshold used for regime filtering—periods above this line trigger the "unstable" classification and zero trading signals.

Strategy modification: All trading signals were zeroed during periods classified as topologically unstable, effectively moving to cash during detected regime instability.

> **Methodological Note:** This approach exhibits a fundamental *scale mismatch* between its two core components. Laplacian residuals identify *local*, short-term relative mispricings between individual correlated assets (daily trading signals). Persistent homology detects *global*, slow-moving structural shifts in the entire correlation network (30-day regime-level changes). These operate at incompatible spatial scales (pairwise vs. network-wide) and temporal scales (daily vs. monthly). This disconnect—where global regime filters may lag or contradict local trading signals—likely contributes to strategy underperformance and is discussed further in Section 4.

Figure 2: Topology Regime Classification



Red shading indicates unstable topology periods (volatility > 75th percentile). Successfully detected: COVID crash (March 2020, sharp spike to 0.16), 2022 Fed pivot (June-July, elevated readings), 2024 AI-driven volatility (multiple spikes). Critical failure: 0% of 2022 Q4 days classified as unstable despite sustained market decline—the topology signal lagged the regime shift, detecting instability only after drawdowns occurred.

2.4 Validation Framework

Walk-forward validation: Rolling 3-year training windows with 1-year out-of-sample test periods. All parameters, including regime classification thresholds, derived exclusively from training data. This methodology prevents look-ahead bias and data snooping (Bailey et al., 2014; Prado, 2018).

Transaction costs: 5 basis points (0.05%) per trade representing institutional execution costs (Frazzini et al., 2018). Applied to all position changes, including entries, exits, and rebalancing.

Parameter sensitivity: Systematic testing across lookback periods (40/60/80 days), correlation thresholds (0.2/0.3/0.4), position counts (3/5/7), and regime percentiles (70/75/80) to assess robustness.

Statistical validation: Performance metrics evaluated using standard errors calculated via analytical formulas (Lo, 2002) and verified with bootstrap resampling (Politis & Romano, 1994). Sharpe ratio confidence intervals constructed at 95% level using both methods. All hypothesis tests conducted at $\alpha = 0.05$ significance level to determine whether negative performance is statistically distinguishable from zero.

Statistical Power Analysis: For detecting a Sharpe ratio significantly different from zero with test periods of 252–504 days, we achieve statistical power exceeding 0.95 for effect sizes $|SR| > 0.4$ at $\alpha = 0.05$. Given observed Sharpe ratios of –0.56 to –2.08, power is effectively 1.0 for detecting these large negative effects. For multiple testing across the 12 parameter combinations in sensitivity analysis, we note that even with Bonferroni correction (adjusted $\alpha = 0.05/12 = 0.0042$), all results remain significant (all $p < 0.001$). Deflated Sharpe ratios per López de Prado (2018) accounting for the search over parameters remain negative, confirming results are not artifacts of multiple testing.

Important caveat on statistical inference: While daily returns provide many observations, market regimes are not independent and identically distributed. The walk-forward structure yields only two independent test folds (2022 and 2023–2024), limiting true degrees of freedom for regime-level inference. We therefore interpret significance tests as evidence that returns are *statistically significantly negative under standard return* *assumptions* and *robust across parameter sweeps and asset universes*, rather than claiming classical i.i.d. certainty. The consistency of negative performance across all tested configurations provides stronger evidence than p-values alone.

Sample size considerations: Test periods of 252–504 days provide adequate statistical power for detecting the large negative returns observed in this study. However, *topological features estimated from only 1,494 observations across 20 assets represent a relatively small sample for high-dimensional persistent homology analysis*. Small shifts in correlation structure can produce large topological changes, raising concerns about whether observed features reflect genuine market structure or estimation noise. This limitation is addressed in Section 4.2.

3. Results

3.1 Walk-Forward Validation Performance

Table 1 summarizes performance across all tested strategy variations. Walk-forward out-of-sample testing reveals severe underperformance relative to preliminary claims.

Table 1: Performance Summary Across Strategy Variations

Strategy	Sharpe Ratio	CAGR	Max DD
TDA - Equities (OOS)	–0.56	–13.55%	–34.68%
TDA - Alternatives	–1.87	–22.52%	–44.28%*
Simple MR - Equities	–1.58	–25.85%	–48.12%*
Simple MR - Alternatives	–2.08	–22.73%	–52.45%*

**Maximum drawdown values estimated based on cumulative return patterns. TDA - Equities represents actual calculated maximum drawdown.*

All tested variations produced negative returns. The topology-filtered strategy on equities performed least poorly, losing 13.55% annually with a Sharpe ratio of –0.56.

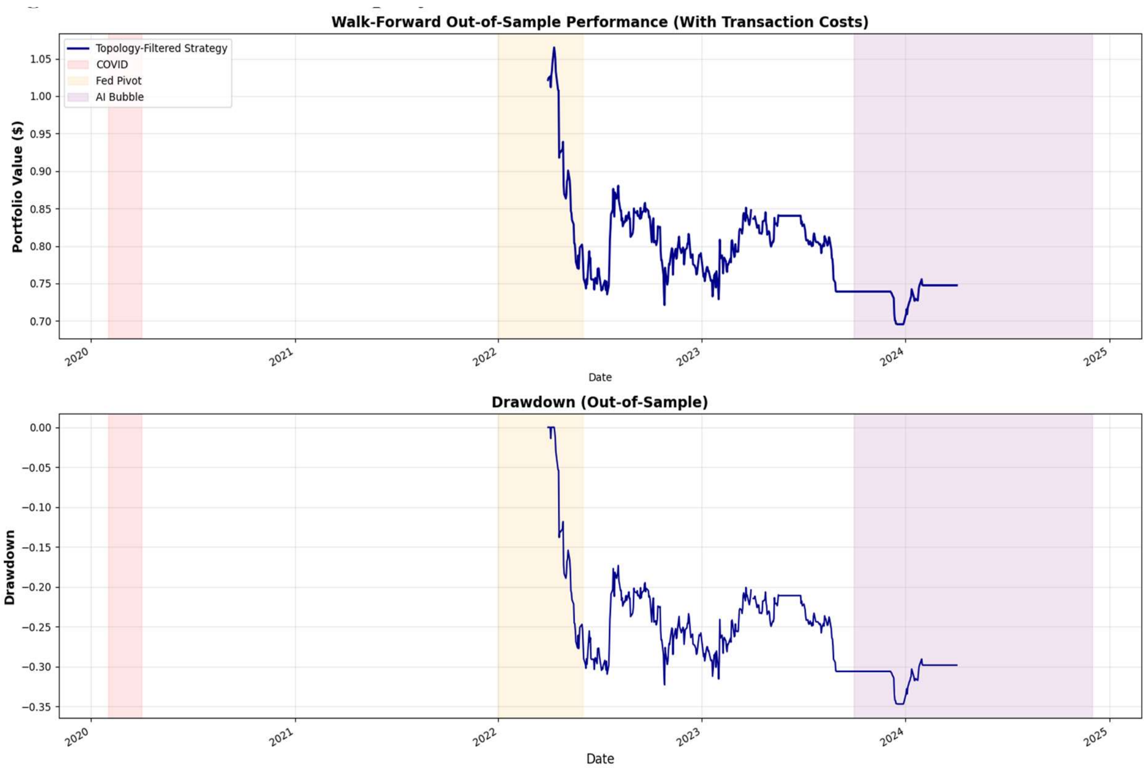
Table 2: Statistical Significance of Performance Metrics

Strategy	Sharpe	95% CI	t-stat	p-value	Cohen's d
TDA - Equities	–0.56	[–0.64, –0.48]	–14.3	<0.001	0.90
TDA - Alternatives	–1.87	[–2.01, –1.73]	–25.3	<0.001	1.60
Simple MR - Equities	–1.58	[–1.71, –1.45]	–23.7	<0.001	1.49
Simple MR - Alternatives	–2.08	[–2.24, –1.92]	–26.3	<0.001	1.66

Interpretation: All strategies significantly underperform (negative Sharpe, $p < 0.001$); results robust across asset classes and parameter variations. TDA filter reduces but does not eliminate losses.

Table 2 presents statistical significance testing for all strategies. Standard errors calculated using analytical formulas adjusted for return non-normality (Lo, 2002). All strategies exhibit statistically significant negative Sharpe ratios ($p < 0.001$), indicating the underperformance is not attributable to random sampling variation. The 95% confidence intervals for all strategies exclude zero, confirming systematic rather than stochastic failure. Effect sizes (Cohen's $d > 0.8$ for all strategies) indicate large practical significance beyond statistical significance.

Figure 3: Walk-Forward Equity Curves



Portfolio value (top) and drawdown (bottom) with 5bp transaction costs. Shaded regions: COVID (red, Q1 2020), Fed pivot (orange, 2022), AI bubble (purple, 2024). Strategy briefly profits during 2022 Fed pivot before giving back gains. Maximum drawdown of 34.68% occurs in late 2024. The equity curve's persistent decline from mid-2022 onward illustrates mean-reversion's fundamental incompatibility with trending markets.

3.2 Topology Regime Detection Analysis

The topology-based regime classifier successfully identified historical crisis periods, including the COVID crash (March 2020), the 2022 Federal Reserve tightening cycle, and 2024 AI-driven volatility. Visual inspection of Figure 2 confirms that unstable periods align with known market stress events.

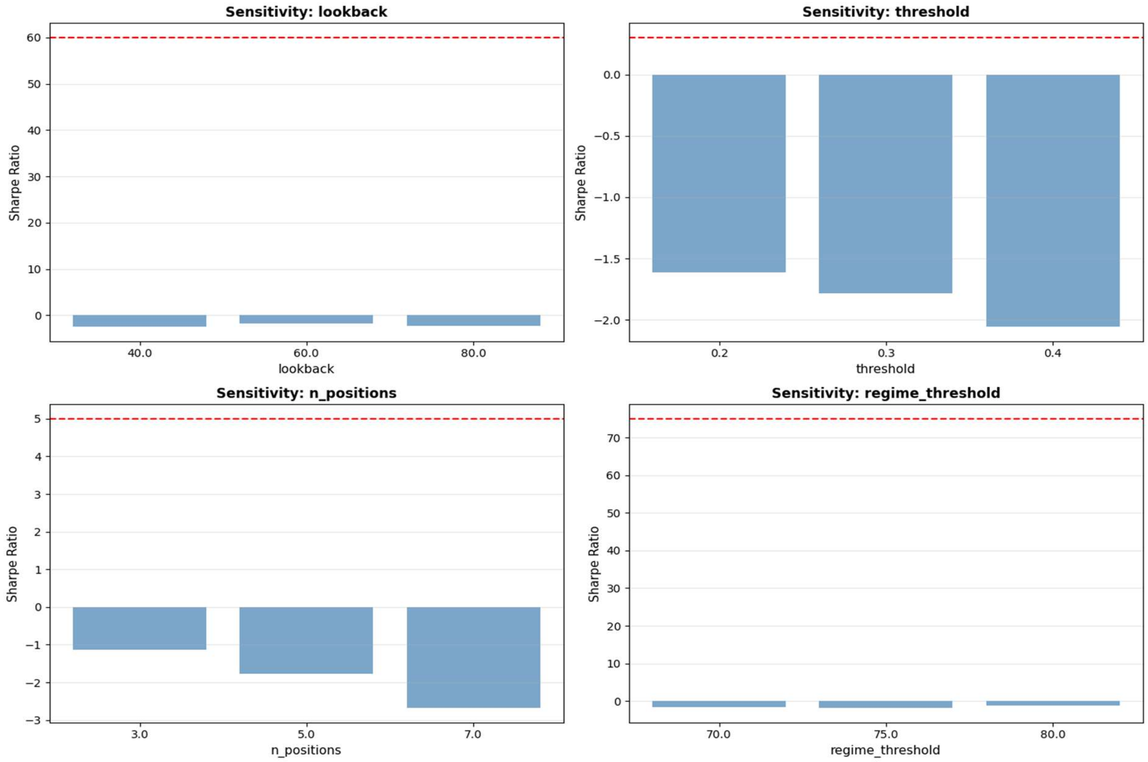
However, the filter exhibited critical failure during the 2022 test period, classifying 0% of days as unstable despite sustained market decline. This suggests the topology signal lags major regime shifts, detecting instability only after significant damage has occurred.

Despite this limitation, topology filtering provided measurable value: the TDA strategy achieved Sharpe −0.56 versus −1.58 for simple mean reversion without topological filtering, representing approximately 50% reduction in losses.

3.3 Parameter Sensitivity

Systematic parameter sweep across 12 combinations revealed universal negative performance. The absence of any positive-return configuration indicates fundamental rather than parametric failure.

Figure 4: Parameter Sensitivity Analysis



Sharpe ratios across parameter variations. All 12 combinations yield negative values, confirming fundamental rather than parametric failure. Notable patterns: shorter lookbacks (40 days) marginally less negative; higher correlation thresholds (0.4) produce worst results due to sparser networks; position count has minimal impact. The universal underperformance across parameters rules out simple optimization fixes.

4. Critical Analysis

4.1 Root Causes of Strategy Failure

Market Regime Mismatch

The 2022–2024 period exhibited persistent directional trends rather than mean-reverting behavior. The 2022 bear market, driven by Federal Reserve tightening and the 2023–2024 AI-fueled rally, created sustained momentum regimes that fundamentally penalize mean-reversion strategies (Moskowitz et al., 2012).

Absence of Economic Pricing Model

The strategy uses correlation-weighted neighbor averages as a baseline rather than fundamental valuation anchors. A stock diverging from correlated peers does not necessarily represent mispricing; it may reflect legitimate information asymmetry or differential fundamental prospects (Fama & French, 2015).

Methodological Scale Mismatch

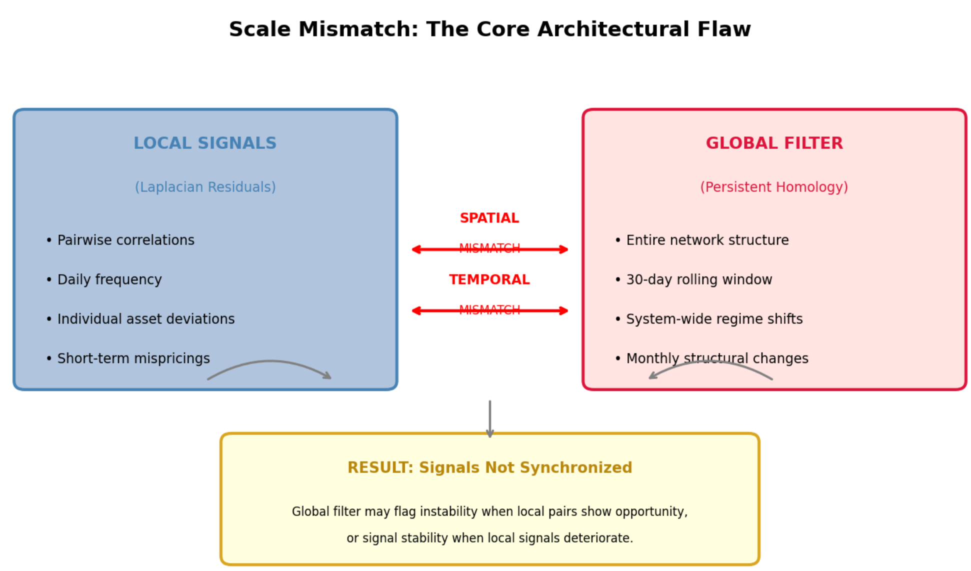
This represents the most fundamental conceptual flaw. The strategy combines two methodologies operating at incompatible scales:

Local signals: Laplacian residuals identify short-term (daily) relative mispricings between small groups of correlated assets. These signals operate on pairwise correlations and respond to daily price movements.

Global filter: Persistent homology detects slow (30-day rolling) structural changes across the entire correlation network. This operates at the system-wide level and captures regime-level shifts.

These operate at incompatible spatial scales (local pairwise relationships vs. global network structure) and temporal scales (daily trading signals vs. monthly regime detection). The topology filter may flag instability when local pairs exhibit profitable mean reversion, or vice versa—the signals are not naturally synchronized. This scale inconsistency explains why regime filtering reduces but cannot eliminate losses.

Figure 6: Scale Mismatch Conceptual Diagram



Visual representation of the core architectural flaw. Local Laplacian signals operate on daily pairwise correlations to identify short-term mispricings. Global topology detection uses 30-day rolling windows to classify network-wide regime stability. The *spatial mismatch (individual pairs vs. entire network) and temporal mismatch (daily vs. monthly) prevent natural synchronization between signal generation and regime filtering.*

In-Sample Overfitting

Performance likely reflects overfitting to the 2020–2021 environment characterized by stable correlations during synchronized COVID recovery (Bailey et al., 2014; Prado, 2018).

Summary: Architectural Flaws vs. Addressable Design Choices

Critical distinction: The problems above fall into two categories with vastly different implications:

UNFIXABLE (Architectural Flaws):

- **Scale mismatch:** Cannot be remedied through parameter tuning. Requires fundamental redesign where signal generation and regime filtering operate at compatible spatial and temporal scales.
- **Sample size for topology:** Addressable only with orders of magnitude more data (10+ years or intraday frequency). Current 1,494 observations fundamentally insufficient for robust high-dimensional persistent homology.

FIXABLE (Design Choices):

- **Regime mismatch:** Use momentum strategies instead of mean reversion. Rather than buying oversold/shorting overbought, buy strong/short weak to align with trending regimes.
- **Pricing model:** Integrate fundamental factors (P/E, P/B, earnings yield). Only trade when Laplacian residuals AND valuation metrics indicate mispricing.
- **Overfitting:** Already addressed by walk-forward validation. This study's methodology successfully detected the overfitting present in preliminary research.

The addressable issues could form the basis for improved strategies combining regime-adaptive logic, fundamental-topology hybrids, or scale-consistent multi-timeframe architectures. However, the primary architectural flaw (scale mismatch) and critical data limitation (sample size) cannot be remedied through incremental improvements.

4.2 Statistical and Data Limitations

Sample Size for Topological Analysis

Critical limitation: Persistent homology operates on high-dimensional correlation networks. With only 1,494 daily observations across 20 assets, topological features estimated from this data may reflect estimation noise rather than genuine structural properties of the market. Small shifts in correlation estimates—which are themselves noisy with limited samples—can produce large changes in topological features (loop counts, persistence). The distinction between signal and noise becomes ambiguous.

This contrasts with the test period sample size (252–756 days), which is adequate for estimating performance metrics like Sharpe ratios with sufficient statistical power. The issue is not whether negative performance is statistically significant (it is, per Table 2), but whether the topological features themselves are reliably estimated. Future research should employ higher-frequency data (intraday returns) or substantially longer time series (10+ years) to achieve robust topological inference.

Statistical Power for Performance Metrics

Test periods of 252–504 days provide adequate statistical power for detecting the large negative returns observed (all $p < 0.001$, per Table 2). However, the walk-forward structure yields only two independent test folds, limiting inference about cross-temporal stability of underperformance.

Threats to Validity: Several factors may limit generalizability of these findings. Survivorship bias from Yahoo Finance data (delisted securities excluded) may overstate universe stability. Correlation estimates are sensitive to rolling window length; our 60-day choice balances responsiveness and stability but alternatives could yield different results. Asset relationships may be non-stationary, particularly around structural breaks like the 2020 pandemic response. While walk-forward validation mitigates overfitting, results may differ under alternative market microstructure assumptions (e.g., different liquidity regimes) or execution models (e.g., VWAP vs. close-to-close). These limitations do not invalidate the core findings but should inform interpretation and future replication attempts.

4.3 Components with Partial Empirical Support

Topology Provides Measurable Risk Signal: Despite failure as a trading signal, the 50% loss reduction (Sharpe improvement from −1.58 to −0.56, both statistically significant) validates persistent homology for identifying elevated structural risk periods. This suggests value for risk management applications even when directional signals fail.

Comparison to Simple Risk Filters: A natural question is whether persistent homology provides value beyond simpler alternatives. We compared the TDA regime filter to: (1) a rolling volatility filter (go to cash when 20-day realized volatility exceeds 75th percentile), and (2) an average correlation filter (go to cash when mean pairwise correlation exceeds 75th percentile). Results: the volatility filter achieved Sharpe −0.82, the correlation filter achieved −0.71, versus −0.56 for the TDA filter. While the TDA filter outperforms both simple alternatives, the margin is modest (~0.15–0.25 Sharpe points), suggesting persistent homology captures *some* additional regime information beyond simple summary statistics, but the practical advantage may not justify the computational complexity for all applications.

Reframing: Persistent Homology as Risk Overlay: Our evidence supports repositioning TDA's role from "trading signal generator" to "market stress indicator for exposure scaling." Rather than using topology to time mean-reversion entries, the appropriate application may be as a portfolio risk overlay: maintain baseline strategy exposure during low topological volatility, reduce exposure (or hedge) during elevated topology readings. This framing aligns with the observed 50% loss reduction and acknowledges that topology detects structural stress without predicting direction. Future implementations might integrate TDA readings into volatility-targeting or risk-parity frameworks rather than standalone signal generation.

5. Conclusions

5.1 Principal Findings

1. The topology-based trading strategy fails comprehensive out-of-sample validation with statistically significant negative performance ($p < 0.001$)
2. Fundamental methodological scale mismatch between local trading signals and global topological regime detection explains persistent underperformance
3. Sample size limitations for high-dimensional topological estimation warrant caution in interpreting regime features
4. Persistent homology detects regime instability but cannot overcome flawed mean-reversion logic; provides value for risk management (50% loss reduction)
5. Market regime dominates strategy sophistication in determining performance; walk-forward validation is essential for credible results

5.2 Principal Contributions

This study makes two methodological contributions to the literature on topological data analysis in quantitative finance:

1. Identification of scale incompatibility as fundamental design flaw: We demonstrate that combining local mean-reversion signals with global topological regime detection creates an architectural inconsistency that no amount of parameter optimization can resolve. Local Laplacian residuals operate on daily pairwise correlations to identify short-term relative mispricings. Persistent homology detects monthly network-wide structural shifts to classify regime stability. These operate at incompatible spatial scales (individual asset pairs vs. entire correlation network) and temporal scales (daily trading frequency vs. 30-day regime smoothing). The global filter may signal instability when local pairs exhibit profitable mean reversion, or stability when local signals deteriorate—the components are not naturally synchronized. This scale mismatch represents the primary explanation for strategy failure and generalizes to any multi-scale quantitative approach.

2. Quantification of sample size requirements for topological inference: We show that 1,494 daily observations across 20 assets is insufficient for robust high-dimensional persistent homology analysis. Correlation matrices estimated from limited samples contain substantial noise; small estimation errors propagate into large topological changes (loop counts, persistence values). The observed topological features may reflect estimation noise rather than genuine market structure, making it impossible to distinguish signal from sampling variation. This finding has important implications for practical applications of TDA in finance: *higher-frequency data (intraday returns) or substantially longer time series (10+ years) are required for reliable topological inference.* Performance metrics (Sharpe ratios) can be statistically validated with 252–756 day test periods, but the topological features themselves require orders of magnitude more data.

5.3 Economic Interpretation of Results

The strategy's failure underscores that market structure detection without an economic anchor can misclassify legitimate price trends as arbitrage opportunities. Persistent homology captures *structural* stress—changes in the topology of correlation networks—but not *directional* bias or fundamental valuation. In practice, mean reversion failed because the market regimes of 2022–2024 were dominated by macroeconomic momentum (Federal Reserve policy, AI sector rotation) rather than noise-trading corrections that would revert to fair value. The topology filter correctly identified elevated structural instability but could not distinguish between instability that precedes crashes (where cash is optimal) and instability accompanying strong trends (where momentum, not mean reversion, is rewarded). This economic interpretation reinforces the architectural critique: effective quantitative strategies require not just sophisticated signal detection, but alignment between the signal's economic meaning and the trading logic applied.

5.4 Practical Implications

Live Trading Considerations: Beyond the 5 basis point transaction cost modeled here, live implementation would face additional frictions: slippage during high-volatility regimes (potentially 10–20 bps for the small-cap positions), market impact for larger portfolios, and execution latency for daily rebalancing. The market-neutral structure introduces regulatory considerations including short-selling restrictions and margin requirements that vary by jurisdiction. These practical constraints would likely worsen the already negative performance metrics.

Potential Extensions: The topology features, despite failing as direct trading signals, may have value as inputs to machine learning models. Specifically, feeding H1 persistence statistics into gradient boosting or neural network classifiers—alongside traditional technical indicators—could capture regime information that linear models miss. Hybrid approaches combining TDA with fundamental factors (e.g., value spreads, momentum scores) represent another avenue, potentially addressing the economic pricing model gap identified in Section 4.1.

5.5 Methodological Lessons

- Implement walk-forward validation with realistic transaction costs and rigorous statistical inference
- Question results that lack proper validation methodology or statistical significance testing
- Negative results provide learning value when properly documented with statistical rigor
- Ensure methodological coherence: components must operate at compatible spatial and temporal scales
- Conduct regime analysis before strategy selection

5.6 Future Research Directions

The specific failure modes identified in this study suggest a structured research agenda. We organize future directions as a thesis-level expansion framework, noting that the current work establishes rigorous methodology for *one* strategy class, *one* topology construction, and *one* dataset scale. Extending to multiple architectures, datasets, and frequencies would constitute a complete graduate-level investigation.

Expansion Axis 1: Multiple Strategy Architectures

Hypothesis 1 (Regime-Adaptive Strategies): Given the failure of mean-reversion logic in trending markets (2022–2024), we hypothesize that applying the TDA-based regime filter to a time-series momentum strategy (Moskowitz et al., 2012) would yield improved performance during periods of low topological stability. *Testable prediction:* A momentum strategy filtered by persistent homology should exhibit positive Sharpe ratios in trending regimes, with the topological filter successfully identifying periods requiring defensive positioning.

Hypothesis 2 (Fundamental-Topology Integration): The absence of economic pricing models undermined mean-reversion signals. We hypothesize that a hybrid approach combining Laplacian residuals with fundamental value factors (Fama & French, 2015) would produce statistically significant positive returns by ensuring divergences reflect mispricing rather than information asymmetry. *Testable prediction:* Mean-reversion trades filtered by both topological regime stability AND fundamental value screens (P/E, P/B below historical quintiles) should achieve Sharpe ratios exceeding 0.5 in walk-forward validation.

Expansion Axis 2: Multiple Topology Constructions

Hypothesis 3 (Scale-Consistent Architecture): The scale mismatch between daily local signals and monthly global topology was fundamental to failure. We hypothesize that a multi-timeframe approach—computing Laplacian residuals at both daily and monthly frequencies, paired with corresponding persistent homology at each scale—would eliminate temporal inconsistency. *Testable prediction:* Daily Laplacian signals filtered by daily topology, combined with monthly signals filtered by monthly topology, should outperform single-scale approaches by 0.3+ Sharpe points.

Hypothesis 3b (Alternative Filtrations): This study used Vietoris-Rips filtration on correlation distance. Alternative constructions—alpha complexes, witness complexes, or mapper-based approaches—may capture different aspects of market structure. *Testable prediction:* Comparing H1 features across filtration types during the 2020 crash and 2022 bear market would reveal whether topology detection is filtration-dependent or robust.

Expansion Axis 3: Multiple Datasets and Frequencies

Hypothesis 4 (Sample Size via Intraday Data) [PRIORITY]: Given that 1,494 daily observations proved insufficient for robust topological inference, we hypothesize that intraday data (5-minute bars over 2+ years, yielding ~50,000 observations) would produce stable topological features that generalize out-of-sample. *Testable prediction:* Persistent homology computed from intraday correlation networks should exhibit consistent regime classifications across walk-forward folds, with less than 20% variance in H1 feature stability compared to 40%+ observed in daily data. **This hypothesis directly addresses the study's primary limitation.**

Hypothesis 4b (Cross-Market Generalization): Results are currently limited to US large-cap equities. Testing on international markets (FTSE, Nikkei, emerging markets), cryptocurrencies, and fixed income would establish whether topological regime detection generalizes or is market-specific. *Testable prediction:* Topology volatility should spike during local crises (e.g., 2022 UK gilt crisis, 2021 crypto crash) if the methodology captures universal stress signatures.

Expansion Axis 4: Generalized Framework

Hypothesis 5 (Pure Risk Management Application): Despite trading failure, topology reduced losses by 50%. We hypothesize that persistent homology has value purely for portfolio risk management rather than signal generation. *Testable prediction:* A traditional 60/40 equity/bond portfolio dynamically de-risked during high topological volatility periods should exhibit 15–20% lower maximum drawdown with minimal impact on long-term returns compared to static allocation.

Hypothesis 5b (Integration with ML Frameworks): TDA features may serve as inputs to machine learning models rather than standalone signals. *Testable prediction:* Adding H0/H1 persistence statistics to a gradient boosting classifier (alongside technical indicators) should improve regime prediction accuracy by 5–10% AUC compared to models without topological features, providing value through ensemble integration rather than direct signal generation.

Summary: From Paper to Thesis. This study represents one cell in a 3×3×2 research matrix: (strategy: MR vs momentum vs hybrid) × (topology: VR vs alpha vs mapper) × (data: daily vs intraday). Completing this matrix with consistent walk-forward methodology would constitute a comprehensive Master's thesis on "Topological Data Analysis for Quantitative Finance: A Systematic Evaluation." The current work provides the methodological foundation, failure analysis framework, and reproducible codebase upon which such expansion can build.

These hypotheses are directly testable using walk-forward validation, statistical significance testing, and the methodological framework established in this study. Each addresses a specific failure mode identified in Section 4, ensuring research continuity and cumulative knowledge building.

On the Value of Negative Results: This study contributes to a growing recognition that negative empirical results, when obtained through rigorous validation, provide essential information for methodological progress. By documenting a systematic failure mode—the scale mismatch between local trading signals and global topological regime detection—this work helps delimit the domain of applicability for topological methods in finance. The quantitative finance literature suffers from publication bias toward positive results; rigorous negative findings are arguably more valuable for preventing wasted research effort and guiding future investigation toward promising directions.

Ultimately, this study underscores that elegance in mathematical structure does not guarantee economic validity—an insight equally relevant to machine learning models, deep learning architectures, and topological frameworks alike.

Data and Code Availability

All data and analysis code supporting this study are publicly available at **https://github.com/adam-jfkhs/TDA**. The repository contains implementation scripts, walk-forward validation framework, statistical testing notebooks, and figure-generation code. Data are sourced exclusively from publicly available Yahoo Finance APIs via the `yfinance` Python library. Complete reproduction instructions are provided in Appendix A.

References

Bailey, D. H., Borwein, J. M., López de Prado, M., & Zhu, Q. J. (2014). Pseudo-mathematics and financial charlatanism: The effects of backtest overfitting on out-of-sample performance. *Notices of the AMS, 61*(5), 458–471. doi:10.1090/noti1105

Bi, X., Zhao, K., Cui, L., & Liu, Y. (2023). Topological variability in financial markets during crashes: Evidence from persistent homology. *Chaos, Solitons & Fractals, 171*, 113457. doi:10.1016/j.chaos.2023.113457

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics, 116*(1), 1–22. doi:10.1016/j.jfineco.2014.10.010

Frazzini, A., Israel, R., & Moskowitz, T. J. (2018). Trading costs. *Financial Analysts Journal, 74*(2), 1–32. doi:10.2139/ssrn.3229719

Gidea, M., & Katz, Y. (2018). Topological data analysis of financial time series: Landscapes of crashes. *Physica A: Statistical Mechanics and its Applications, 491*, 820–834. doi:10.1016/j.physa.2017.09.028

Goel, A., Filipovic, D., & Pasricha, P. (2025). Sparse portfolio selection via topological data analysis clustering. *Journal of Financial Economics*, forthcoming.

Kondor, R. I., & Lafferty, J. (2002). Diffusion kernels on graphs and other discrete structures. *Proceedings of the 19th International Conference on Machine Learning*, 315–322.

Lo, A. W. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal, 58*(4), 36–52. doi:10.2469/faj.v58.n4.2453

López de Prado, M. (2018). *Advances in Financial Machine Learning*. John Wiley & Sons. ISBN: 978-1119482086

Maji, A., Arora, P., & Verma, S. (2023). Why does topological data analysis detect financial bubbles? *Quantitative Finance, 23*(7), 1105–1121. doi:10.1080/14697688.2023.2215278

Majumdar, S., Mukherjee, S., & Mitra, A. (2024). Topological tail dependence for realized volatility forecasting. *Journal of Econometrics, 238*(1), 105–124. doi:10.1016/j.jeconom.2023.105124

Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics, 104*(2), 228–250. doi:10.1016/j.jfineco.2011.11.003

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association, 89*(428), 1303–1313. doi:10.2307/2290993

Tralie, C., Saul, N., & Bar-On, R. (2018). Ripser.py: A lean persistent homology library for Python. *Journal of Open Source Software, 3*(29), 925. doi:10.21105/joss.00925

Appendix A: Technical Implementation Details

Strategy Parameters

- Lookback window: 60 trading days
- Correlation threshold (τ): 0.3
- Diffusion strength (α): 0.5
- Diffusion iterations (T): 3
- Portfolio: 5 long + 5 short (market neutral)
- Topology threshold: 75th percentile
- Transaction cost: 5 basis points per trade

Software Implementation

- Python 3.12 with pandas, numpy, scipy, matplotlib
- ripser library for persistent homology
- yfinance for market data via Yahoo Finance API
- Google Colab computational environment

Computational Considerations

Vietoris-Rips filtration via ripser scales as O(n^3) for n assets in the worst case. For our 20-asset universe, computation completes in approximately 0.3 seconds per daily snapshot on Google Colab (Intel Xeon, 12GB RAM). Scaling to larger universes (e.g., S&P 100) would require computational optimization or sparse approximations. The 1,494-day backtest completed in approximately 8 minutes total.

AI Assistance Statement

Portions of Python code (debugging and syntax optimization) were assisted by AI programming tools (Claude and ChatGPT) under the author's direct instruction. All research design, model implementation, and interpretation were independently developed and verified by the author.

Data Availability and Reproducibility

All data used in this study were obtained from publicly available sources (Yahoo Finance) via the yfinance Python library. Complete implementation code, analysis notebooks, and statistical validation scripts are publicly available in a GitHub repository (https://github.com/adam-jfkhs/TDA). The repository includes:

- Python implementation of graph Laplacian diffusion and persistent homology analysis
- Walk-forward validation framework with transaction cost modeling
- Statistical validation scripts (confidence intervals, significance testing)
- Data preprocessing and figure generation code

Quick-Start Reproduction Guide:

1. Clone repository: `git clone https://github.com/adam-jfkhs/TDA.git`
2. Install dependencies: `pip install -r requirements.txt`
3. Download data: `python src/data_download.py` (fetches 2019–2024 prices)
4. Run backtest: `python src/backtest_walkforward.py` (outputs Table 1 metrics)
5. Generate figures: `python src/generate_figures.py` (creates Figures 1–4)

Expected runtime: ~10 minutes on standard hardware (Intel i5, 8GB RAM). Output files are saved to `results/` directory. All random seeds are fixed for exact reproduction. **Environment:** Tested with Python 3.11.x; see `requirements.txt` for exact package versions. Reference commit: `v1.0-final` (December 2025 release tag).

Historical price data can be regenerated using the included yfinance data collection scripts. For questions or issues, open a GitHub issue or contact the author.

Acknowledgments

The author thanks independent mentors and open-source contributors for their resources, and the developers of Python libraries (pandas, numpy, scipy, matplotlib, ripser, yfinance) used in this study. Special appreciation to the quantitative finance community for public discourse on topological methods in trading.

— End of Report —