

7476038-persistent-homology-in-an-emerging-equity-market-dependence-structure-regime-sta

AVTORJI: Raziye Rafibakhsh, MohammadAmin Rezazadeh

POVZETEK: Raziskava obravnava, ali je persistentna homologija informativna pri opisovanju zavisnosti in tržnih regij v rastujočem trgu akcij, ter ali jo to dodatno vlagajo v predviđanje prihodnje realizirane variјanc. Raziskovalci analizirajo 2.769 dnevnih opazovanj petih indeksov Tehreanskega trga akcij od 26. marca 2014 do 22. septembra 2025. Nastopajoči rezultati pokažejo, da tradicionalne indikatorji vlagajo značilno, vendar nekompletno v topološko raznolikost. Topologija se razlikuje, če se trenutno zavisnost odstrani, zato so Gaussova kovariјansa in Gaussova kopulna surrogatna metode boljši v reprodukciji topološkega razpada. Analiza regij je identificirala stabilen granicni čas v junu 2018 in manj stabilen v aprilu 2022. Topologija se najboljše opisuje z okno velikosti 50 dneh, pri katerih dodate topologije podatke zgoraj se vlagajo v predviđanje, vendar s podmnožico natančnosti. Nelinearna učenje zmanjša izbiro natančnosti, vendar je njeno interval vključno z nuljo in predviđa prihodnje realizirane variјanci le pod določenimi pogoji.

METODA: Raziskovalci uporabljajo 2.769 dnevnih opazovanj petih Tehreanskega trga akcij indeksov, ki se razlikujejo v 50 dnevnih oknehu, z 30 in 100 dnevnim oknom za odvisnost. Uporabljajo štiri merila traje in konvencionalne indikatorje volatilnosti in zavisnosti. Surrogatne metode uporabljajo za ustrezno statistično analizo. Regij analizirajo brez uporabe topologije in ocenjujejo prihodnje realizirane variјanci z vloženim validacijo in nepotrđljivim ohranilcem.

KLJUČNE UGOTOVITVE:

- Tradicionalni indikatorji vlagajo značilno, vendar nekompletno v topološko raznolikost.
- Odstranitev trenutne zavisnosti vpliva veliko na topologijo.
- Gaussova kovariјansa in Gaussova kopulna surrogatna metode reproducuje večjo distribucijo topologije.
- Regija analiza identificira stabilen granicni čas v junu 2018 in manj stabilen v aprilu 2022.
- Dodate topologije podatke zgoraj se vlagajo v predviđanje, vendar s podmnožico natančnosti.
- Nelinearna učenje zmanjša izbiro natančnosti, vendar je njeno interval vključno z nuljo in predviđa prihodnje realizirane variјanci le pod določenimi pogoji.

POMEN ZA TRGOVANJE: Topologija je najboljše opisana z oknom velikosti 50 dneh, pri katerih dodate topologije podatke zgoraj se vlagajo v predviđanje. Topologija je najbolj informativna kot kompaktne nelinearne predstave zavisnosti in regij. Vlaganje topologije v predviđanje zmanjša natančnost, vendar je njeno vlaganje v predviđanje prihodnje realizirane variјanci pod določenimi pogoji koristno.

Persistent Homology in an Emerging Equity Market: Dependence Structure, Regime Stability, and Realized Variance Forecasting

Raziye Rafibakhsh*

MohammadAmin Rezazadeh[†]

September 2026

Abstract

Market stress simultaneously changes volatility, dependence, and common risk concentration, creating a need for multidimensional representations of evolving market structure. We therefore examine whether persistent homology provides an informative representation of dependence structure and market regimes in an emerging equity market, and whether it adds incremental predictive value for future realized variance. We analyze 2,769 daily observations for five Tehran Stock Exchange indices from 26 March 2014 to 22 September 2025. Rolling point clouds use a primary window of 50 trading days, with windows of 30 and 100 trading days used for robustness. We compare four persistence measures with conventional indicators of volatility and dependence, investigate their statistical origins with three surrogate mechanisms, characterize regimes without using topology, and evaluate forecasts of realized variance over the next 20 trading days with nested validation and an untouched holdout. Classical indicators explain a meaningful but incomplete share of topological variation. Removing contemporaneous dependence produces large departures from the observed topology, whereas Gaussian covariance and Gaussian copula surrogates reproduce substantially more of its distribution. Regime analysis identifies a stable boundary in June 2018 and a less stable boundary in April 2022. With the window fixed at 50 days for both variable sets, adding topological data analysis variables worsens holdout QLIKE for ridge regression. A nonlinear learner improves selected holdout losses, but its QLIKE interval includes zero and it exceeds a naive persistence forecast only under QLIKE. Fixed windows of 30 and 100 days also yield no reliable QLIKE improvement. Persistent homology is therefore most informative as a compact nonlinear representation of dependence concentration and regime evolution. Its forecasting contribution remains conditional on the learner and evaluation criterion.

Keywords: persistent homology; topological data analysis; dependence structure; market regimes; volatility forecasting; Tehran Stock Exchange; surrogate data **JEL Codes:** C58; C63; G10; G17

1 Introduction

Episodes of market stress typically involve simultaneous changes in volatility, dependence among assets, and the concentration of common risk. Standard indicators such as realized volatility, average

*Affiliation of the first author. Email: Rbr.razi@email.com

[†]Affiliation of the second author. Email: aminrezazadeh21@email.com

correlation, and the leading eigenvalue of the correlation matrix summarize these channels through prespecified moments or linear dependence. Topological data analysis (TDA) provides a complementary description of the geometry of multivariate observations across scales. Persistent homology, in particular, tracks geometric features as a proximity threshold varies and converts their birth and death scales into numerical summaries suitable for statistical analysis (Carlsson G , 2009; Edelsbrunner H, Harer J , 2010).

Methods based on sliding windows and time delay embeddings provide a foundation for applying persistent homology to observations ordered in time (Perea JA, Harer J , 2015). In finance, Gidea M, Katz Y (2018) documented pronounced movements in persistence landscape norms around major crashes in United States equity markets. Aguilar A, Ensor KB (2020) developed comparisons based on mean persistence landscapes, while Guo H, et al. (2020) examined the 2008 global financial crisis and the European debt crisis using indices from the United States, Europe, and China. Related work used persistent homology to forecast adverse market regimes (Baitinger E, Flegel S , 2021) and combined persistence measures with diagnostics for critical slowing down (Ismail MS, et al. , 2022). Evidence is not uniform across samples. Valdivia AD (2023) found that changes in several persistence summaries were informative around historical downturns but did not establish a universal rule for early warning. More recent studies offered an explanation based on the log periodic power law singularity model for why topological signals may arise during financial bubbles (Akingbade SW, et al. , 2024). Other work incorporated features obtained from TDA into forecasts produced by neural networks (de Jesus Jr LC, et al. , 2025). Taken together, this literature supports the responsiveness of topological summaries to changes in market organization, but it does not imply that every topological movement provides an early warning signal or contains incremental predictive information.

The value of TDA also depends on the task for which it is used. Using prices for 60 European firms, Hobbelhagen F, Diamantis I (2024) compared TDA with an extended symbolic aggregate approximation (SAX). TDA responded more strongly to localized fluctuations, whereas SAX represented broader movements more clearly. This result suggests that a complex representation should be evaluated relative to its intended financial task and relevant benchmarks. For applications in market risk, the important questions are therefore whether topological changes are distinct from familiar volatility and dependence effects, whether they remain stable across alternative specifications, and whether they are useful in a genuinely forward evaluation.

Across the studies reviewed above, three empirical issues remain insufficiently connected. First, persistence summaries are affected by the scale and dependence structure of the underlying point cloud, but their overlap with classical indicators of market conditions is not always quantified directly. Second, relatively little evidence identifies which distributional properties of returns generate the observed topology by comparing the data with surrogates that selectively preserve marginal distributions, covariance, or copula dependence. Third, alignment between a topological statistic and a crisis date within the estimation sample does not by itself establish a stable regime break or incremental forecasting value. Separating these claims requires sensitivity analysis across specifications for rolling windows and change point detection, followed by a comparison outside the estimation sample with strong classical baselines and controls that prevent information leakage.

We study these questions using 2,769 daily observations for five Tehran Stock Exchange (TSE) indices from 26 March 2014 to 22 September 2025: TEDPIX, the Second Market Index, the Equal

Weighted Index, the Chemicals Index, and the Top 30 Index. The five return series exhibit substantial common variation. Their pairwise correlations range from 0.672 to 0.963, and the first eigenvalue accounts for 88.7% of the variation in the correlation matrix. At the same time, the indices differ in their weighting, coverage, or sectoral focus. This concentrated index system provides a demanding setting in which to test whether topological summaries capture information beyond volatility and linear dependence. We construct rolling point clouds with windows of 30, 50, and 100 trading days. We then summarize homology in dimension one using norms of persistence landscapes, persistent entropy, feature counts, and maximum persistence.

The empirical design separates four research questions that are often conflated. RQ1 asks whether topological features respond to changing market conditions. RQ2 asks to what extent their information is distinct from conventional measures of volatility and dependence. RQ3 asks how stable the identified market regimes are across alternative specifications. RQ4 asks whether topological features improve forecasts of future realized variance after conventional predictors have been included. We first assess stability across window lengths and explain topological variation using volatility, dependence, market mode concentration, effective rank, and indicators based on distance. We then compare the observed features with surrogates that preserve marginal independence, Gaussian covariance, or a Gaussian copula. Change points are evaluated across window lengths and alternative detection specifications. Finally, we test whether topology adds value to forecasts of realized market variance over 20 days. The forecasting procedure uses nested validation with an expanding sample, an embargo matched to the forecast horizon, inference based on a block bootstrap, and a final holdout sample that remains untouched during model selection.

This study makes three contributions. First, it combines classical explanatory benchmarks with targeted surrogate experiments to attribute financial topology to identifiable distributional features rather than only documenting movements around selected events. Second, it distinguishes robust evidence of market regimes from breaks that are sensitive to specification choices by evaluating change points across alternative rolling windows and detection designs. Third, it evaluates the incremental value of TDA in a strictly forward setting with linear and nonlinear learners. This integrated design supports a more precise assessment of where persistent homology adds information and where conventional indicators of market conditions absorb its apparent value.

The results support a qualified interpretation. For the primary specification with a window of 50 trading days, classical indicators of market conditions explain between 20.9% and 32.3% of the adjusted variation in the reported topological summaries. Destroying dependence among the index returns produces substantial departures from the observed topology, whereas surrogates that preserve covariance or copula dependence reproduce considerably more of its behavior. The regime analysis identifies a highly stable break on 23 June 2018 and a weaker break on 12 April 2022 that is more sensitive to specification choices. In the final forecasting holdout, adding the topological variables to the primary ridge model increases QLIKE, and the bootstrap interval indicates deterioration rather than improvement. The nonlinear model improves selected error measures, but its gain in QLIKE is statistically inconclusive, and the naive forecast of realized variance remains competitive. Analyses with fixed windows of 30 and 100 trading days also produce no reliable QLIKE gain. The evidence is therefore stronger for TDA as a tool for monitoring market structure and characterizing regimes than as an input that uniformly improves forecasts.

The remainder of the paper is organized as follows. Section 2 presents the data and methodology. Sections 3, 4, and 5 report the empirical results. Section 3 examines the information content and structural origins of the topological features. Section 4 evaluates regime characterization and stability. Section 5 evaluates incremental forecasting performance outside the estimation sample. Section 6 discusses the interpretation and limitations, and Section 7 concludes.

2 Data and methodology

2.1 Data and return construction

The empirical analysis uses daily observations for five Tehran Stock Exchange (TSE) indices: the Tehran Price Index (TEDPIX), the Second Market Index, the Equal Weighted Index, the Chemicals Index, and the Top 30 Index. The indices were selected to represent the aggregate market, a distinct market segment, an alternative weighting scheme, a major sector, and large capitalization firms. This composition provides a multivariate view of the market while retaining meaningful differences in index coverage and construction. Historical index levels were retrieved from the TSE data services using FinPy-TSE version 1.2.10, an open source Python package for programmatic data access (Rahimi A, Ghafouri R, 2024). The final sample contains 2,769 trading day observations from 26 March 2014 to 22 September 2025.

Let $P_{i,t}$ denote the closing level of index i on trading day t . Continuously compounded returns are defined as

$$r_{i,t} = \ln(P_{i,t}) - \ln(P_{i,t-1}).$$

Dates were converted to a common trading day index, arranged chronologically, and checked for duplicates, missing values, and nonfinite observations. No duplicate dates, missing returns, or infinite values remained in the validated panel. As an additional source check, all TEDPIX returns were independently reconstructed from the underlying price series. The maximum absolute difference between the supplied and reconstructed returns was 1.79×10^{-15} , which is within numerical precision. All rolling statistics are dated at the final observation of their respective windows, so a value recorded at t uses information no later than t .

The unconditional correlations among the five return series are high, ranging from approximately 0.67 to 0.96. The first eigenvalue of the full sample correlation matrix accounts for 88.7% of its trace. This strong common component makes the panel a demanding setting for the analysis: a topological feature must be evaluated against the possibility that it primarily reflects familiar marketwide dependence rather than an independent source of information.

2.2 Rolling point cloud representation

For each trading day t and window length w , the multivariate return window is written as

$$\mathcal{X}_{t,w} = \{\mathbf{r}_s \in \mathbb{R}^5 : s = t - w + 1, \dots, t\},$$

where $\mathbf{r}_s = (r_{1,s}, \dots, r_{5,s})$ is the return vector in five dimensions. We consider $w \in \{30, 50, 100\}$. The primary specification uses 50 trading days. Windows of 30 and 100 trading days provide checks

over shorter and longer horizons.

Because the indices have different marginal scales, each coordinate is standardized separately within every window. If the sample mean and standard deviation of coordinate j in $\mathcal{X}_{t,w}$ are $\mu_{j,t,w}$ and $s_{j,t,w}$, the point cloud coordinates are

$$z_{j,s} = \frac{r_{j,s} - \mu_{j,t,w}}{s_{j,t,w}}.$$

The resulting w observations form a standardized point cloud $\mathcal{Z}_{t,w}$ in $\mathbb{R}^5$. Standardization within each window removes differences in location and marginal variance. The analysis therefore focuses on the relative geometry of joint returns and their dependence structure. The transformation does not remove nonlinear dependence or changes in the shape of the joint distribution. A scale tolerance of 10^{-12} provides a numerical safeguard. The code records the status of every window, so failed calculations cannot be removed without documentation.

2.3 Persistent homology features

Persistent homology describes how geometric structures evolve as the distance threshold used to connect observations increases (Edelsbrunner H, Harer J, 2010). For each standardized point cloud, we construct a Vietoris Rips filtration using Euclidean distance. At threshold ϵ , two observations are connected when their distance does not exceed ϵ . A simplex is added when every pair of its vertices is connected. The filtration is not truncated at a prespecified finite distance. We study homology in dimension one, denoted by H_1 , which records loops in the point cloud.

Each H_1 feature is represented by a birth and death pair (b_k, d_k) , with persistence length $\ell_k = d_k - b_k$. Only finite pairs satisfying $d_k > b_k$ are retained. From the collection of persistence lengths in each window, we compute the number of valid H_1 features, n_{H_1} ; total, mean, median, and maximum persistence; persistent entropy; and the L_1 and L_2 norms of the complete persistence landscape. Persistent entropy is

$$\text{PE} = - \sum_{k=1}^n p_k \ln p_k, \quad p_k = \frac{\ell_k}{\sum_{j=1}^n \ell_j}.$$

For the complete persistence landscape, the norms are obtained directly from persistence lengths (Bubenik P, 2015):

$$\|\Lambda\|_1 = \frac{1}{4} \sum_{k=1}^n \ell_k^2,$$

$$\|\Lambda\|_2 = \left(\frac{1}{12} \sum_{k=1}^n \ell_k^3 \right)^{1/2}.$$

These expressions calculate the norms without an arbitrary evaluation grid. When a window contains no valid H_1 interval, summaries for which zero is mathematically defined are set to zero. Feature construction produced 2,740, 2,720, and 2,670 observations for windows of 30, 50, and 100 trading days, respectively. No calculation failed, and no infinite H_1 interval was present. The window of 30 trading days produced 38 empty H_1 diagrams, whereas the two longer windows produced none.

Eight summaries were initially retained for feature auditing: n_{H_1} , total persistence, the two landscape norms, persistent entropy, maximum persistence, mean persistence, and median persistence. Because several of these quantities are algebraically or empirically redundant, the principal structural and forecasting analyses emphasize four representatives: the L_2 landscape norm, persistent entropy, n_{H_1} , and maximum persistence. The wider set is retained for descriptive information content checks.

2.4 Classical market state benchmarks and information content

To determine whether the topological measures contain information beyond conventional summaries, eight classical features are calculated from the same raw return window. Average marginal volatility is the mean of the five sample standard deviations. Market mode volatility is the square root of the largest eigenvalue of the covariance matrix. Mean pairwise correlation is the average of the ten distinct correlations outside the diagonal. Dependence concentration is measured by the largest eigenvalue of the correlation matrix divided by its trace. Effective correlation rank is defined as

$$\text{ER} = \exp\left(-\sum_{j=1}^5 q_j \ln q_j\right),$$

where q_j is eigenvalue j of the correlation matrix divided by the sum of its nonnegative eigenvalues. A smaller effective rank indicates that variation is concentrated in fewer common dimensions. We also calculate the proportion of covariance variation explained by the first principal component. Finally, the mean and coefficient of variation of all pairwise Euclidean distances are calculated after standardization within each window. The distance benchmarks are therefore expressed in the same geometry used for persistent homology.

The association between each topological feature $T_{t,w}$ and the classical vector $\mathbf{B}_{t,w}$ is assessed using Pearson and Spearman correlations, mutual information with five nearest neighbours, and the linear explanatory model

$$T_{t,w} = \alpha_w + \mathbf{B}_{t,w}' \boldsymbol{\beta}_w + u_{t,w}.$$

The principal estimand is the model R^2 , interpreted as the fraction of variation in a topological series statistically explained by the classical feature set. It is not interpreted as proof that the two representations are equivalent. To account partially for the serial dependence induced by overlapping windows, coefficient uncertainty is estimated with heteroskedasticity and autocorrelation consistent covariance matrices using $w-1$ lags (Newey WK, West KD, 1987). Predictor variables are standardized before coefficient comparison, and the values of p for individual coefficients are adjusted within each specification using the Benjamini and Hochberg procedure (Benjamini Y, Hochberg Y, 1995). Because the classical predictors are correlated, individual coefficients receive secondary emphasis relative to the joint explanatory power of the model.

2.5 Surrogate data tests

The regression analysis measures statistical overlap but does not identify which properties of the joint return distribution generate the observed topology. We therefore compare the empirical persistence features with three surrogate mechanisms. Each surrogate is generated from the raw observations

within a given rolling window and is subsequently standardized and processed through exactly the same persistent homology pipeline as the observed data.

The Gaussian covariance surrogate draws observations from a multivariate normal distribution with the empirical mean vector and covariance matrix of the window. It asks whether Gaussian structure defined by the first two moments is sufficient to reproduce the observed feature. The marginal independence surrogate permutes the observations separately within each index. It preserves the empirical marginal distribution of every index but removes contemporaneous dependence among indices. For the Gaussian copula surrogate, ranks are converted to normal scores and their correlation matrix is estimated. New observations are then drawn from a multivariate normal distribution with that correlation matrix. This construction preserves the fitted Gaussian copula dependence but imposes Gaussian marginal distributions. Comparisons across these null models separate the roles of empirical marginals, linear covariance, and dependence represented by the fitted copula.

For every evaluated window and surrogate type, 199 replications are generated from seed 42. To limit computation while retaining temporal coverage, the series based on 50 trading days is evaluated every five observations. The series based on 30 and 100 trading days are evaluated every 20 observations. The final observation is added when it does not fall on this grid. This design produces 818 evaluated windows and 9,816 comparisons between four features and three surrogate types. For an observed feature value T and surrogate values $T^{(1)}, \dots, T^{(M)}$, we report the surrogate mean, a standardized deviation, the fraction of observations outside the pointwise 95% surrogate envelope, and a two sided empirical value of p . The code calculates the latter from the smaller empirical tail and applies a finite sample correction:

$$p = \min \left\{ 1, 2 \min \left(\frac{1 + \sum_{m=1}^M \mathbb{I}(T^{(m)} \leq T)}{M + 1}, \frac{1 + \sum_{m=1}^M \mathbb{I}(T^{(m)} \geq T)}{M + 1} \right) \right\}.$$

Results adjusted by the Benjamini and Hochberg procedure are also reported. These window specific tests are treated as descriptive because adjacent rolling windows remain dependent; correction for multiple tests does not eliminate this temporal dependence. Moreover, rejection of a Gaussian null with matched covariance is interpreted as evidence that the first two Gaussian moments are insufficient. It is not automatically interpreted as economically useful information about higher moments.

2.6 Independent regime characterization

Market regimes are estimated independently of the topological variables. This separation prevents regimes from being defined by the same TDA features later used to describe them. Six rolling classical indicators are employed: average marginal volatility, market mode volatility, mean pairwise correlation, the share represented by the largest correlation eigenvalue, effective correlation rank, and the maximum TEDPIX drawdown within the window. The primary calculation uses 50 trading days, and calculations based on 30 and 100 trading days assess robustness.

Each indicator is robustly standardized, and changes in its distribution are identified with the Pruned Exact Linear Time (PELT) algorithm using a radial basis function cost (Killick R, et al. , 2012). The primary specification imposes a minimum segment length of 120 observations, evaluates candidate locations in increments of five observations, and uses penalties $c \ln(n)$, with $c \in \{3, 4, 5, 6, 8\}$. Candidate breaks occurring within 20 trading observations are grouped into clusters. A primary

consensus break must be supported by at least four of the six indicators and at least three penalty values, while accepted consensus breaks must be separated by at least 120 observations.

Sensitivity is evaluated through four additional designs. Two designs use windows of 30 and 100 trading days. A relaxed design uses a minimum segment of 90 observations and lower penalties, while a conservative design uses a minimum segment of 150 observations and higher penalties. Breaks are matched uniquely across specifications when their locations differ by no more than 30 observations. The recovery rate of each primary break and pairwise agreement between specifications quantify stability. Regime labels are descriptive rather than predictive. A classical stress score is constructed from the indicators after robust standardization. The sign of effective rank is reversed so that a larger value always indicates greater stress. Regimes are then described using the lower and upper terciles of this score. All scaling and PELT estimation in this section use the complete sample. The regimes are therefore treated as ex post structural descriptions and are never used as a prediction target or a real time trading signal.

2.7 Incremental out of sample forecasting design

The final analysis asks whether TDA improves forecasts of future TEDPIX variance after classical information has been included. The target observed at date t is the annualized variance over the following 20 trading days:

$$RV_{t,t+20} = 252 \times \frac{1}{20} \sum_{h=1}^{20} r_{\text{TEDPIX},t+h}^2.$$

Every predictor is available at t , whereas the target ends at $t + 20$. The common forecasting panel contains 2,650 dates from 24 August 2014 to 20 August 2025 for each candidate window length. The baseline combines TEDPIX realized variance over 5, 20, and 60 days, the absolute current return, and the eight classical variables defined above. The realized variance terms are motivated by heterogeneous autoregressive models (Corsi F, 2009). The augmented specification adds the L_2 landscape norm, persistent entropy, n_{H_1} , and maximum persistence. Incremental value is measured by comparing the same algorithm with and without the TDA variables, rather than by comparing unrelated model classes.

Ridge regression is the primary forecasting model because shrinkage addresses the strong dependence among predictors and provides a transparent test of incremental linear information. The dependent variable is the logarithm of future variance. Forecasts are transformed back to the variance scale and bounded below by 10^{-12} . The ridge penalty is selected from $\{0.1, 1, 10, 100\}$. A histogram based gradient boosting regressor provides a nonlinear robustness check. Its learning rate is 0.05, the maximum number of iterations is 300, and the minimum leaf size is 30. The tuning grid combines maximum leaf counts of $\{7, 15\}$ with L_2 penalties of $\{1, 10\}$.

The most recent 20% of eligible dates, comprising 530 observations from 12 June 2023 to 20 August 2025, is reserved as an untouched final holdout. The development sample contains 2,100 dates and ends on 9 May 2023. Twenty intervening dates form an embargo so that no target in the development sample overlaps the holdout period. Evaluation within the development sample uses five outer folds with an expanding training sample. The confirmatory comparison fixes the rolling window at 50 trading days for both variable sets. Within every outer training period, three additional folds with the same expanding structure select only the model parameters by validation QLIKE. The

complete procedure is repeated with fixed windows of 30 and 100 trading days as robustness checks. A secondary analysis allows the inner folds to select the window separately for each variable set; this analysis evaluates two complete forecasting procedures and is not interpreted as a pure test of the contribution from TDA variables. A gap of 20 observations separates each training block from the corresponding validation or test block. Standardization and parameter selection therefore use only observations available within each training fold.

Table 1 summarizes the forecasting sample split, validation structure, and controls used to prevent information leakage.

Table 1: Forecasting design and controls against information leakage.

Design element	Specification
Market	Five index panel from the Tehran Stock Exchange
Return observations	2,769 trading days
Forecast ready dates	2,650 common dates
Forecast evaluation period	24 August 2014 to 20 August 2025
Rolling windows	50 trading days in the primary comparison; fixed windows of 30 and 100 days for robustness
Forecast target	Annualized realized variance over the next 20 trading days
Development sample	2,100 dates
Gap before the holdout	20 observations
Untouched holdout	530 dates, from 12 June 2023 to 20 August 2025
Validation folds	5 outer folds and 3 inner folds with expanding training periods; inner folds select model parameters
Primary model and loss	Ridge regression and QLIKE
Nonlinear robustness model	Histogram gradient boosting
Bootstrap inference	1,999 circular block resamples with a block length of 50

The primary loss is QLIKE (Patton AJ , 2011),

$$\text{QLIKE}_t = \ln(\widehat{v}_t) + \frac{v_t}{\widehat{v}_t},$$

where v_t and $\widehat{v}_t$ are realized and predicted future variance. Lower values indicate better forecasts. Robustness metrics are mean absolute error and root mean squared error on the volatility scale, together with root mean squared error of log variance. For each observation, incremental improvement equals the loss of the baseline model minus the loss of the same model with TDA variables. Positive values therefore favor the augmented specification.

Uncertainty for the final holdout improvement is estimated using a paired circular block bootstrap with 1,999 resamples and a block length of 50 observations (Künsch HR , 1989). The same resampled indices are applied to the loss sequences from both feature sets, preserving their paired nature and part of the serial dependence generated by overlapping forecast horizons. We report the mean improvement, the 95% percentile confidence interval, and the bootstrap probability that the improvement is positive. Ridge and QLIKE form the confirmatory combination of model and metric. Gradient boosting and the alternative losses are interpreted as robustness evidence. The ex post regimes from Section 2.6 do not enter this forecasting design.

2.8 Reproducibility and interpretation protocol

All major specifications are held constant across modules: a primary window of 50 trading days, alternative windows of 30 and 100 trading days, persistent homology in dimension one, standard score normalization within each window, a forecast horizon of 20 trading days, and random seed 42. Intermediate datasets retain window start and end dates, specification identifiers, quality control fields, and model selections. This permits every result to be traced to the corresponding input window and prevents failed calculations from being removed without documentation.

The empirical sequence is deliberately hierarchical. Feature auditing establishes numerical validity and redundancy; classical comparisons assess statistical overlap; surrogate tests investigate structural origins; independent change point analysis evaluates descriptive regime stability; and the final forward validation exercise tests predictive value. Evidence at one stage is not treated as proof at another. In particular, a topological feature may differ from a surrogate null or coincide with a market transition without providing incremental out of sample forecasting value.

3 Information content and structural origins of topological features

This section evaluates whether the persistent homology features are numerically reliable, stable across rolling horizons, and distinct from conventional descriptions of market volatility and dependence. The analysis proceeds in three stages. We first examine the distribution, persistence, and stability of the topological summaries across window lengths. We then measure how much of their variation can be explained by classical market state variables. Finally, experiments with surrogate data identify the contribution of contemporaneous dependence among indices to the observed topology.

3.1 Feature validity and temporal behavior

The construction procedure produced 8,130 combinations of date and specification: 2,740 for the window of 30 trading days, 2,720 for the window of 50 trading days, and 2,670 for the window of 100 trading days. No calculation failed, and no infinite H_1 interval was retained. Empty H_1 diagrams occurred in 38 observations under the shortest window, corresponding to 1.39% of that series. No empty diagram occurred under either longer window. This complete accounting matters because removing windows with little topological structure could bias the series toward turbulent periods.

Table S1 in the Supplementary Material summarizes the four features used in the principal structural and forecasting analyses. As expected, the number of detected loops increases strongly with the number of points in the cloud: mean n_{H_1} rises from 3.83 at 30 days to 8.76 at 50 days and 24.21 at 100 days. Persistent entropy shows a similar mechanical response because longer windows contain more intervals over which persistence can be distributed. The landscape L_2 norm and maximum persistence also increase with window length, although less sharply. Differences in feature levels across window lengths should therefore not be interpreted as differences in economic stress. Movements within a given specification, standardized comparisons, and rank agreement on common dates are more informative.

All four series display substantial first order persistence. With a window of 50 trading days, lag one autocorrelations range from 0.910 for maximum persistence to approximately 0.928 for the landscape norm and n_{H_1} . Persistence becomes stronger with a window of 100 trading days. This pattern is expected because consecutive estimates share $w - 1$ observations. It should not be read as independent

evidence of slowly evolving economic states. It does, however, justify inference that accounts for overlap and the use of blocked resampling in subsequent analyses.

The eight summaries considered during the initial audit also exhibit considerable redundancy. With a window of 50 trading days, three feature pairs have an absolute Pearson correlation above 0.90. Six pairs exceed 0.90 when the larger absolute value from the Pearson and Spearman correlations is used. The first three principal components explain 96.9% of total standardized variation, while the first component alone accounts for 67.4%. Similar patterns are obtained with windows of 30 and 100 trading days. The analysis therefore does not treat every persistence statistic as a separate discovery. The landscape L_2 norm, entropy, loop count, and maximum persistence are retained as representatives of aggregate lifetime magnitude, lifetime dispersion, topological complexity, and the most prominent individual feature.

3.2 Stability across rolling window specifications

Agreement across window lengths is positive but moderate. Across the eight audited summaries, the median absolute Spearman correlation is 0.496 between windows of 30 and 50 trading days and 0.468 between windows of 50 and 100 trading days. It falls to 0.356 when windows of 30 and 100 trading days are compared. No pair of specifications has a median absolute Spearman correlation above 0.70. Half of the features exceed 0.50 when the 30 and 50 day windows are compared. The corresponding proportions are 37.5% for the comparison between 50 and 100 days and 12.5% for the comparison between 30 and 100 days. Complete pairwise stability statistics for the eight audited features are reported in Table S7 of the Supplementary Material.

These findings have two implications. First, the consistently positive rank relationships indicate a common component across horizons. Second, the choice of window materially changes the empirical signal. Short windows respond more rapidly and occasionally generate empty diagrams, whereas long windows produce smoother and more persistent series. The window of 50 trading days therefore provides a compromise between responsiveness and stability, but its results are not invariant to temporal scale. Estimates based on 30 and 100 trading days are retained as genuine robustness checks rather than interchangeable replications.

3.3 Relationship with classical market state indicators

Table 2 reports adjusted R^2 values from regressions of each topological feature on the eight classical market state indicators. With the primary window of 50 trading days, the classical variables explain between 20.9% and 32.3% of feature variation. The largest adjusted R^2 is obtained for the landscape L_2 norm (0.323), followed by n_{H_1} (0.318), mean persistence (0.303), entropy (0.266), maximum persistence (0.254), and median persistence (0.209). Conventional measures of volatility, dependence, concentration, and distance therefore account for a meaningful portion of each topological series. A contemporaneous linear combination of these benchmarks nevertheless leaves most of the variation unexplained.

The explanatory share generally rises with the rolling horizon. With a window of 100 trading days, adjusted R^2 values range from 0.261 for maximum persistence to 0.508 for mean persistence. This increase is especially visible for n_{H_1} , mean persistence, and median persistence. Smoothing over longer windows appears to strengthen the component shared with slowly moving measures of

Table 2: Structural evidence for the four prespecified topological features at the primary window of 50 trading days.

Panel A: Explanatory fit and strongest classical association

Topological feature	Adjusted R^2	Strongest Pearson association
Landscape L_2 norm (H_1)	0.323	ER (0.524)
Maximum persistence (H_1)	0.254	LER (-0.445)
Number of H_1 features	0.318	PC1 (-0.551)
Persistent entropy (H_1)	0.266	PC1 (-0.489)

Panel B: Surrogate evidence

Topological feature	BH independence	BH covariance	BH copula	Mean z , independence
Landscape L_2 norm (H_1)	61.1%	0.0%	0.0%	-1.957
Maximum persistence (H_1)	45.7%	0.0%	0.0%	-1.694
Number of H_1 features	49.7%	0.0%	0.0%	-2.440
Persistent entropy (H_1)	53.2%	0.0%	0.0%	-3.418

Notes: Adjusted R^2 comes from a regression on the eight classical market state indicators. The reported Pearson association has the largest absolute value among those indicators. LER denotes the share of the largest correlation eigenvalue, PC1 denotes the share of the first principal component, and ER denotes effective correlation rank. BH columns report the percentage of evaluated windows that remain significant after adjustment by the Benjamini and Hochberg procedure. The final column reports the mean standardized deviation under marginal independence. Results for windows of 30 and 100 trading days are treated as robustness evidence.

correlation and concentration. In contrast, the explanatory power of maximum persistence changes less, which suggests that the most prominent loop retains a larger component specific to that feature.

Pairwise associations clarify the direction of this overlap. At 50 days, the landscape L_2 norm is positively correlated with effective correlation rank (Pearson $r = 0.524$) and negatively associated with mean pairwise correlation in ranks (Spearman $\rho = -0.630$). Entropy and n_{H_1} are most strongly negatively correlated with the first principal component share ($r = -0.489$ and -0.551 , respectively). Maximum, mean, and median persistence are also negatively related to correlation concentration or mean correlation. When market returns become more concentrated along a common direction, the standardized cloud becomes effectively lower dimensional and supports fewer or weaker loop structures. Conversely, greater effective dimensionality is associated with richer H_1 topology.

The nonlinear association results do not overturn this interpretation. Mutual information is strongest for benchmarks related to volatility or dependence, although its magnitude varies materially by feature and horizon. Because mutual information is unconditional and the baseline predictors are themselves highly correlated, it cannot identify an independent causal channel. The condition number of the regression design also increases from 267 with a window of 30 trading days to 648 with a window of 100 trading days. This degree of collinearity cautions against attaching structural meaning to individual regression coefficients. The joint R^2 and the consistency of rank relationships therefore provide the primary evidence.

Overall, the classical comparison rejects two extreme interpretations. The topological features are not merely duplicates of a single volatility or correlation statistic, because a substantial residual component remains. However, they are also not detached measures of market complexity: their dominant associations are systematically linked to dependence concentration and effective dimensionality. The relevant question is therefore what aspect of the joint distribution produces the residual topological variation.

3.4 Surrogate evidence on structural origins

The surrogate experiments provide a sharper attribution test. Figure 1 and Table S2 in the Supplementary Material present results for the primary window of 50 trading days. Under marginal independence, the observed features are consistently below their surrogate counterparts. The mean standardized deviation ranges from -1.69 for maximum persistence to -3.42 for persistent entropy. Depending on the feature, 56.0% to 66.1% of observed windows fall outside the individual 95% surrogate envelope, and 45.7% to 61.1% remain significant after adjustment by the Benjamini and Hochberg procedure. Permuting the five return series independently therefore changes the topology substantially.

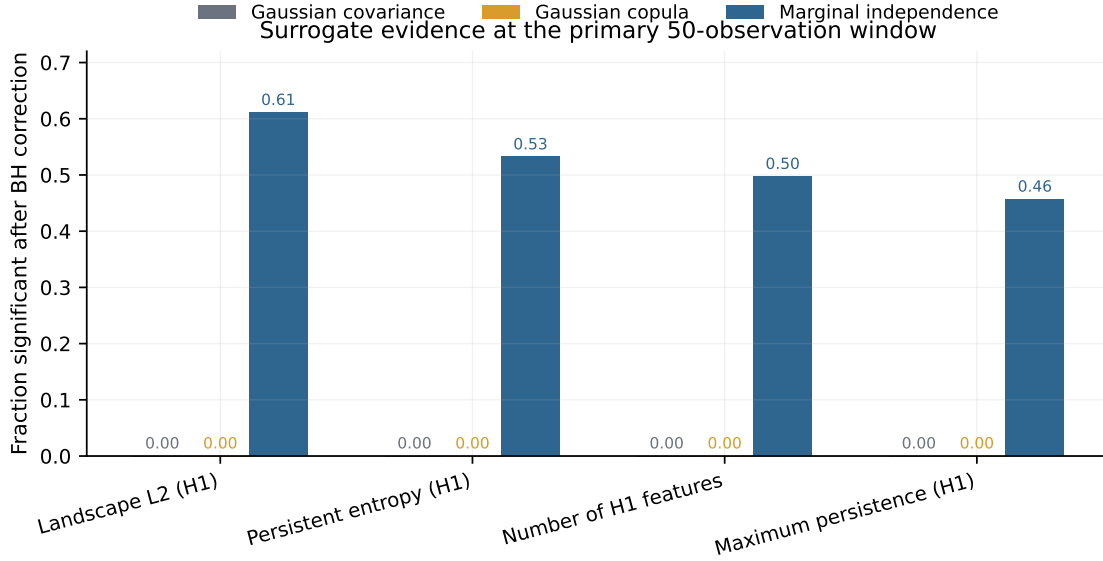


Figure 1: Surrogate evidence for the primary window of 50 trading days. Surrogates with independent marginals depart strongly from the observed topology, whereas surrogates that preserve dependence reproduce substantially more of its distribution.

The direction of the difference is consistent with the dependence mechanism identified above. Surrogates with independent marginals generate more loops, higher entropy, and larger persistence magnitudes than the observed market. Strong contemporaneous common movement concentrates the empirical point cloud along fewer effective directions. This concentration reduces the number and size of H_1 structures relative to clouds assembled from the same marginal returns without alignment among indices. The result complements the negative association between topological features and correlation concentration reported in Section 3.3.

By contrast, the Gaussian covariance and Gaussian copula surrogates reproduce much more of the observed topology. With a window of 50 trading days, their mean standardized deviations remain negative but are substantially closer to zero. The fraction outside the 95% envelope ranges from 7.5% to 14.1% for the covariance surrogates and from 10.5% to 18.2% for the copula surrogates. For both null models, no window remains significant after adjustment by the Benjamini and Hochberg procedure for any feature. The primary evidence therefore does not support a broad claim that the observed H_1 summaries systematically contain structure beyond the dependence captured by these fitted Gaussian models.

The alternative windows reinforce this ordering. Under marginal independence, the fraction that

remains significant after adjustment ranges from 22.5% to 28.3% with a window of 30 trading days, from 45.7% to 61.1% with a window of 50 trading days, and from 50.4% to 86.7% with a window of 100 trading days. In contrast, this fraction is zero for all covariance and copula comparisons at every horizon. Even under these two null models, the fraction outside the unadjusted 95% envelope rises with window length, particularly for n_{H_1} and entropy. Longer windows may therefore expose systematic deviations from the fitted null distributions. These deviations are treated as suggestive because they do not survive adjustment across windows and because adjacent windows remain dependent.

Both dependence preserving surrogates match the observed topology much more closely than the independence surrogate. However, the fitted Gaussian copula does not consistently outperform the Gaussian model with empirical covariance; for several features, the covariance surrogate is at least as close to the observed value. Both mechanisms impose Gaussian marginal distributions, so their comparison cannot identify the contribution of non Gaussian marginals. The strongest and most stable implication is therefore narrower: contemporaneous dependence is essential, but the evidence does not establish a clear advantage for either fitted dependence model.

3.5 Synthesis

Regarding RQ1, the persistent homology features are computationally well behaved and vary systematically with observable measures of market dependence and concentration, but their levels and temporal smoothness depend materially on window length. Across features and window lengths, adjusted R^2 values range from approximately 0.21 to 0.51; for the primary 50 day specification, they range from 0.21 to 0.32. Dependence concentration is the most consistent link: stronger common movement is associated with fewer and weaker loop structures.

Regarding RQ2, the surrogate tests show that the distinction between dependence and independence is considerably more important than the difference between the two fitted Gaussian dependence models. Removing contemporaneous dependence produces large and persistent changes, whereas the covariance and copula surrogates reproduce most of the observed behavior after adjustment for multiple tests. TDA therefore provides a nonlinear summary of the geometry induced by market dependence. Its residual component is not negligible, but the present evidence does not establish that this component is unique to topology. Section 4 examines whether the topological patterns correspond to stable market regimes, and Section 5 evaluates whether they improve forecasts of risk.

4 Independent regime characterization and stability

This section examines whether broad market regimes can be identified without using the topological features themselves. The separation is deliberate: defining regimes from persistence statistics and then using those regimes to validate the same statistics would produce circular evidence. We instead estimate regime boundaries exclusively from six classical indicators of volatility, dependence, dimensional concentration, and drawdown. The analysis first reports the primary segmentation based on a window of 50 trading days, then characterizes the resulting regimes, and finally evaluates the sensitivity of the detected breaks to alternative windows and segmentation thresholds.

4.1 Consensus change points

The primary PELT runs generated 160 candidate breaks across the combinations of indicator and penalty. Grouping candidates whose locations differed by no more than 20 observations produced 19 distinct regions. The predefined consensus rule requires support from at least four of the six indicators and at least three of the five penalty values. Only two regions satisfy both conditions. The first change point is dated 23 June 2018 and the second 12 April 2022. These boundaries divide the 2,720 observations available with the window of 50 trading days into three broad regimes.

Table 3 reports the two consensus change points and summarizes their stability across the alternative segmentation designs.

Table 3: Consensus change points and their stability across segmentation designs.

Panel A: Change points under the primary design

Consensus date	Cluster width	Indicator support	Penalty support	Run support
23 June 2018	15	100.0%	100.0%	96.7%
12 April 2022	5	83.3%	80.0%	40.0%

Panel B: Sensitivity to the segmentation design

Specification	Window	Breaks	Regimes	June 2018 recovered	April 2022 recovered
Primary	50	2	3	Yes	Yes
Relaxed	50	9	10	Yes	Yes
Conservative	50	1	2	Yes	No
Window of 30 days	30	3	4	Yes	Yes
Window of 100 days	100	2	3	Yes	No

Notes: Cluster width is measured in observations. Recovery allows a distance of at most 20 observations from a primary change point. The design with a window of 30 days locates the second boundary 20 observations earlier than the primary design. The design with a window of 100 days locates the first boundary five observations later and does not recover the second boundary.

The 2018 break receives exceptionally strong support within the primary specification. All six indicators and all five penalty settings detect it, and 29 of the 30 possible runs contribute a candidate to the cluster. The contributing locations span 15 observations, which indicates some uncertainty about the precise boundary. Nevertheless, the support rate of 96.7% shows that the transition is not driven by one variable or a narrow choice of penalty.

The 2022 boundary satisfies the acceptance criteria but is less decisive. Five indicators and four penalty settings support it, and it appears in 12 of the 30 runs. TEDPIX drawdown is the only indicator that does not support this cluster. The contributing locations span only five observations, so the detections that do occur are closely aligned. Their frequency across all runs is only 40.0%, however, which indicates weaker overall agreement. The two breaks should therefore not be assigned equal evidential weight.

Candidate breaks outside these two clusters fail at least one consensus condition. Their exclusion is substantively important because individual series and runs with low penalties identify numerous brief changes. The consensus rule is intended to isolate persistent shifts shared across several dimensions of the market rather than produce a chronology of every volatility episode.

4.2 Economic profiles of the regimes

Figure 2 and the regime profiles reported in the Supplementary Material summarize the mean classical indicators within the three regimes. Regime 1 covers 10 June 2014 to 20 June 2018 and contains 975 observations. Its average stress score is -0.835 , well below the values in the later regimes. Average marginal volatility is 0.0053 , market mode volatility is 0.0107 , and mean pairwise correlation is 0.659 . The effective correlation rank of 2.26 is also substantially higher than in the subsequent periods, indicating that market variation is distributed across more independent directions. On this basis, Regime 1 is classified as a low stress state.

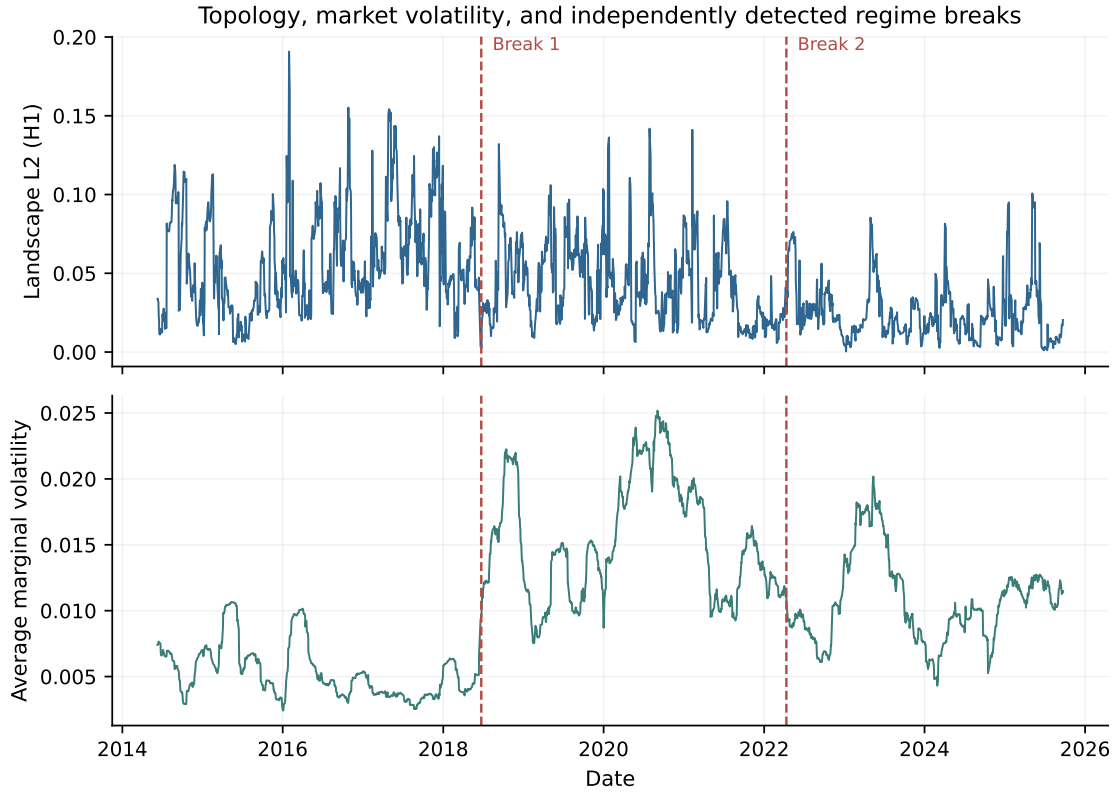


Figure 2: Evolution of the primary landscape norm and market volatility with independently estimated PELT consensus breaks.

Regime 2 begins at the first consensus break and lasts until 11 April 2022, covering 920 observations. All principal stress dimensions intensify. Average marginal volatility rises to 0.0153 , almost three times its Regime 1 value, while market mode volatility increases to 0.0333 . Mean pairwise correlation reaches 0.863 and the largest correlation eigenvalue share rises from 0.736 to 0.893 . At the same time, effective correlation rank falls from 2.26 to 1.53 and mean lookback drawdown increases from 0.051 to 0.127 . The joint movement toward higher volatility, stronger synchronization, greater concentration, and larger drawdowns produces the highest mean stress score, 0.383 . Regime 2 is therefore characterized as high stress.

Regime 3 runs from 12 April 2022 to the end of the sample and contains 825 observations. Its mean stress score declines to 0.230 , leading to a moderate stress classification. Volatility and drawdown recede relative to Regime 2: average marginal volatility falls to 0.0107 , market mode volatility to 0.0233 , and drawdown within the window to 0.096 . Synchronization among indices does not reverse,

however. Mean pairwise correlation rises further to 0.905, the share represented by the leading correlation eigenvalue reaches 0.925, and effective rank falls to 1.38.

This contrast is informative for interpreting the regimes. The period after April 2022 is less volatile than the preceding high stress regime but more concentrated in a common market direction. Regime labels should therefore not be interpreted as a ranking based only on volatility. The composite classification distinguishes the intensity of realized fluctuations from the structure of common movement. Regime 3 represents partial normalization in volatility and drawdown together with persistent, and even stronger, synchronization.

The profiles also provide an independent interpretation of the topological relationships reported in Section 3. Stronger correlation concentration and lower effective rank after 2018 describe a market geometry with fewer effective dimensions. Section 3 shows that such conditions are generally associated with fewer or weaker H_1 structures. The two sets of results are coherent, but the topological variables do not enter the construction of the regime boundaries. The correspondence is therefore interpretive rather than a mechanical confirmation.

4.3 Sensitivity to the segmentation design

The number of detected regimes varies substantially with the degree of segmentation conservatism. The relaxed specification reduces the minimum segment length to 90 observations and requires support from only three indicators. It yields nine consensus breaks and ten regimes. The conservative specification uses a minimum segment length of 150 observations and requires support across four penalty values. It retains one break and produces two regimes. The primary design lies between these extremes, with two breaks and three regimes.

This sensitivity is expected: relaxed settings admit shorter and less widely supported transitions, whereas conservative settings retain only the most persistent common shift. It also shows why selecting a single penalty after inspecting the results would be problematic. The consensus framework makes the degree of support visible and allows the stability of each primary boundary to be evaluated rather than treating the chosen segmentation as uniquely correct.

Changing the rolling window produces fewer additional regimes than relaxing the segmentation thresholds. The specification based on 30 trading days identifies three breaks and four regimes. It recovers both primary breaks, although the match to the second break occurs 20 observations earlier, on 8 March 2022. The specification based on 100 trading days identifies two breaks and three regimes but recovers only the first primary boundary, displaced by five observations to 30 June 2018. Longer windows smooth local transitions and can shift or remove weaker breaks, while shorter windows respond more quickly and admit an additional boundary.

4.4 Stability of individual breaks

The distinction between the two primary breaks becomes clearer when recovery is assessed across all five specifications. Every design recovers the June 2018 boundary, giving a recovery fraction of 100%. Its matched location is exact under the primary, relaxed, conservative, and 30 day designs. With a window of 100 trading days, the match differs by only five observations. The first boundary is therefore robust to both the rolling horizon and the segmentation threshold.

The primary, relaxed, and 30 day specifications recover the April 2022 boundary, whereas the

conservative and 100 day designs do not. Its recovery fraction is therefore 60%. This boundary is detectable under the primary design and under specifications that retain greater responsiveness, but it does not survive greater smoothing or stricter segmentation. The result is consistent with a meaningful but less stable transition.

Pairwise agreement across complete sets of breaks is moderate rather than uniform. Agreement is highest between the primary and 30 day designs, with a symmetric score of 0.833, and equals 0.750 between the primary and conservative specifications. It falls to 0.500 between the primary and 100 day designs because only one of the two primary breaks is matched. Comparisons involving the relaxed specification are mechanically lower because it contains many additional short segments. These results show why recovery should be reported separately for each break rather than summarized only by the number of regimes or a single agreement score.

4.5 Interpretation and scope

Regarding RQ3, the regime analysis provides one strong structural result and one conditional result. The transition centered on June 2018 is supported across indicators, penalties, window lengths, and segmentation thresholds. It can be treated as a stable boundary in the retrospective organization of the sample. Most primary indicators support the April 2022 transition, but its detection depends more strongly on temporal resolution and segmentation choices. It should therefore be reported as a secondary boundary that is sensitive to specification choices, rather than as an equally established break.

These conclusions concern retrospective characterization. Robust standardization, PELT estimation, consensus formation, and state naming use the complete sample. The change points consequently cannot be interpreted as dates on which an investor would necessarily have identified a new regime in real time. Nor are the regime labels used as forecast targets in Section 5. Their purpose is to determine whether broad shifts in conventional market conditions are stable enough to provide an independent context for interpreting the topological series.

Overall, the change point evidence indicates that a single regime does not adequately describe the TSE sample. This statement concerns structural segmentation and should not be interpreted as the result of a formal test of stationarity. The most persistent change occurs in June 2018 and is followed by substantially higher volatility, stronger common movement, fewer effective dimensions, and deeper drawdowns. The period after April 2022 shows partial moderation in volatility but continued concentration of market risk. This asymmetric pattern, in which scale normalizes without a comparable normalization in dependence, motivates the strict forward test in the next section. That analysis asks whether topological information improves forecasts once these classical conditions are already represented.

5 Incremental forecasting value outside the estimation sample

This section examines whether the topological variables improve forecasts of future TEDPIX variance after the models account for conventional volatility and market state information. The comparison is incremental within each algorithm: the baseline variables are held constant and four TDA measures are then added. To isolate this addition, the primary comparison fixes the rolling window at 50 trading

days for both variable sets. The inner validation folds select model parameters but do not select the window. The complete analysis is repeated with fixed windows of 30 and 100 trading days. A separate analysis in which the window is selected within validation is retained as a comparison between complete forecasting procedures, not as the primary test of the contribution from TDA variables. Ridge regression and QLIKE form the confirmatory analysis. Histogram gradient boosting and alternative losses provide robustness evidence.

5.1 Evaluation sample and forward validation design

Aligning the three rolling specifications with the target, which measures realized variance over the next 20 trading days, yields 2,650 common forecast dates from 24 August 2014 to 20 August 2025. The first 2,100 dates constitute the development sample. We reserve the final 530 dates, beginning on 12 June 2023, as an untouched holdout. A gap of 20 observations separates the development and holdout samples because the target windows for those observations extend into the holdout period.

We evaluate the development sample with five outer folds whose training periods expand through time. Each test block contains 316 observations and follows a gap of 20 observations. The training sample therefore grows from 500 observations in the first fold to 1,764 in the fifth. Within every outer training period, three inner folds with expanding training periods select the model parameters. The rolling window remains fixed within each complete analysis. Table 1 summarizes this temporal structure.

This design provides two complementary forms of evidence. The outer folds indicate whether any incremental forecasting gain recurs across historical periods. The holdout measures performance in the most recent part of the sample after the design and tuning rules have been fixed. Since adjacent targets cover overlapping 20 day intervals, daily forecast errors are serially dependent. We therefore do not treat them as independent observations.

5.2 Performance during model development

Under the primary 50 day specification, the mean QLIKE across the five outer folds is -1.968 for baseline ridge regression and -1.934 after adding the TDA variables. The resulting mean improvement is -0.034 , where improvement is defined as baseline loss minus augmented loss. The augmented model has a higher QLIKE in every fold. The fold improvements are -0.075 , -0.037 , -0.029 , -0.001 , and -0.027 . Mean absolute volatility error, root mean squared volatility error, and the error for log variance also increase on average after the TDA variables are added. These development results provide no evidence of a stable incremental linear contribution.

See Tables S3 and S4 in the Supplementary Material.

Histogram gradient boosting produces more favorable average results but little temporal consistency. Its mean QLIKE falls from -1.152 to -1.198 , which gives an average improvement of 0.047 . Each of the four reported metrics favors the augmented model in only two of the five folds. Under QLIKE, the gains occur in folds 2 and 3, while folds 1, 4, and 5 favor the baseline. The result therefore suggests that the nonlinear model can use the topological variables during selected periods, but not that the gain persists through the development sample.

The parameter choices are made separately for the two variable sets because their predictor dimensions and regularization needs differ. Every ridge fit selects $\alpha = 100$. For gradient boosting, the

choices of terminal leaves and the L_2 penalty vary across folds. Separate parameter tuning does not change the information window and is part of evaluating each model fairly after the TDA variables are added.

5.3 Results for the untouched holdout

Table 4 and Figure 3 report results for the final holdout. With the window fixed at 50 days, both ridge specifications select $\alpha = 100$. Adding the TDA variables increases QLIKE from -2.620 to -2.607 . It also increases mean absolute volatility error from 0.0366 to 0.0374 , root mean squared volatility error from 0.0445 to 0.0455 , and the error for log variance from 0.640 to 0.650 . Thus, all four holdout criteria favor the baseline ridge model.

Table 4: Forecast performance in the untouched final holdout.

Model	Variable set	Window	QLIKE	MAE volatility	RMSE volatility	RMSE log variance	QLIKE gain [95% interval]
Ridge regression	Baseline	50	-2.620	0.0366	0.0445	0.640	NA
Ridge regression	Baseline plus TDA	50	-2.607	0.0374	0.0455	0.650	-0.013 [-0.030, -0.001]
Histogram gradient boosting	Baseline	50	-2.520	0.0477	0.0593	0.818	NA
Histogram gradient boosting	Baseline plus TDA	50	-2.547	0.0431	0.0545	0.759	0.027 [-0.009, 0.063]
Naive forecast	Previous 20 day realized variance	NA	-2.516	0.0399	0.0491	0.729	NA

Notes: The holdout contains 530 observations. A lower loss indicates better performance. QLIKE gain is baseline loss minus the loss from the same algorithm with TDA variables, so a positive value favors the TDA specification. Intervals are percentile intervals from 1,999 paired circular block bootstrap resamples with a block length of 50. The naive forecast equals realized variance over the previous 20 trading days and does not select a TDA window. NA indicates that a window or incremental comparison is not applicable to that row.

Untouched holdout forecasts: baseline versus baseline + TDA

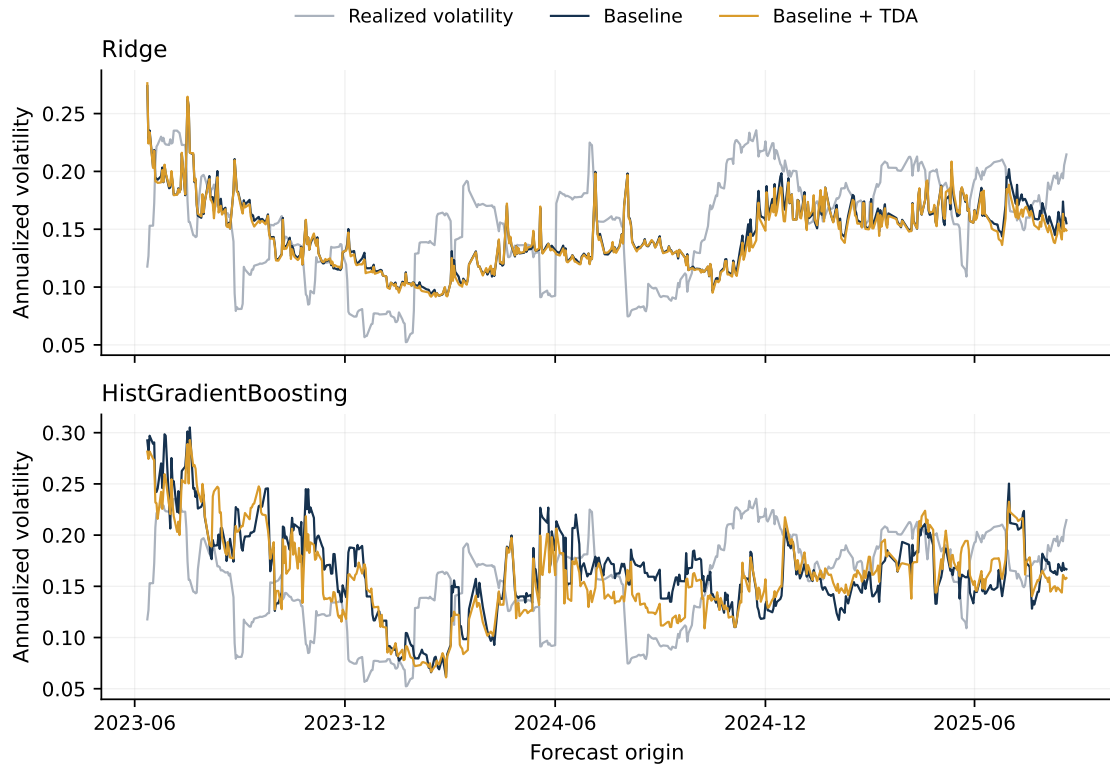


Figure 3: Realized and predicted volatility over the untouched holdout for the linear and nonlinear forecasting models.

For histogram gradient boosting, both variable sets use the same 50 day window and seven terminal leaves. The baseline model uses an L_2 penalty of 1, whereas the augmented model uses a penalty of 10. Adding the TDA variables reduces QLIKE from -2.520 to -2.547 , mean absolute volatility error from 0.0477 to 0.0431 , root mean squared volatility error from 0.0593 to 0.0545 , and the error for log variance from 0.818 to 0.759 .

The comparison with the naive forecast is metric dependent. The gradient boosting model with TDA variables has a lower QLIKE than the forecast based on realized variance over the previous 20 days: -2.547 compared with -2.516 . However, the naive forecast has a lower mean absolute volatility error, a lower root mean squared volatility error, and a lower error for log variance. The corresponding values are 0.0399 versus 0.0431 , 0.0491 versus 0.0545 , and 0.729 versus 0.759 . The nonlinear TDA model therefore does not dominate the naive forecast across evaluation criteria. The baseline ridge model also has a lower holdout QLIKE and a lower root mean squared volatility error than either gradient boosting specification. Any nonlinear improvement should consequently be interpreted within that model class.

5.4 Inference with a block bootstrap

Figure 4 reports results from a paired circular block bootstrap with 1,999 resamples and a block length of 50 observations. For the confirmatory comparison between the two ridge specifications under QLIKE, the mean improvement is -0.0132 and the 95% interval is $[-0.0298, -0.0015]$. The estimated bootstrap probability of a positive improvement is 0.9%. The interval lies below zero, indicating that adding the TDA variables worsens holdout QLIKE under the primary design. The mean ridge differences for absolute and squared volatility errors are also negative, although their intervals include zero.

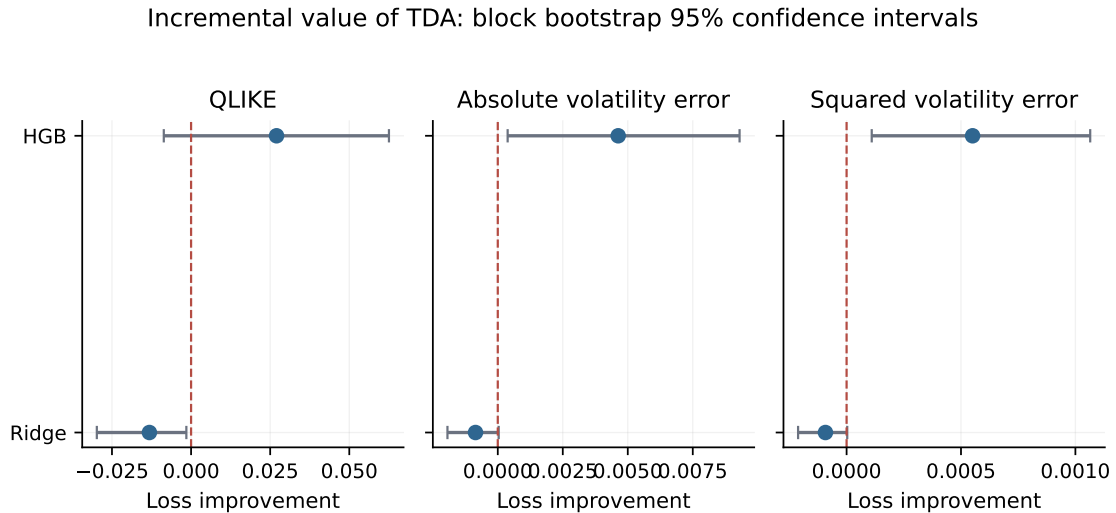


Figure 4: Paired circular block bootstrap confidence intervals for incremental TDA value. Positive values favor baseline plus TDA.

For gradient boosting, the mean QLIKE improvement is 0.027 and its 95% interval is $[-0.009, 0.063]$. This interval includes zero. The mean improvement in absolute volatility error is 0.0046 with an interval from 0.0004 to 0.0093 . The improvement in squared volatility error is 0.00055 with an interval

from 0.00011 to 0.00107. Both intervals exclude zero.

The difference across metrics is plausible because QLIKE evaluates forecasts on the variance scale and penalizes underprediction strongly when realized variance is high. The other two losses measure deviations on the volatility scale. Nevertheless, ridge regression with QLIKE is the designated confirmatory comparison. Results from the nonlinear model and alternative losses are secondary, and no adjustment has been applied for examining several such comparisons. We therefore treat the two intervals that exclude zero as supporting evidence rather than as the main inferential result.

5.5 Robustness across fixed windows

Repeating the complete analysis with each alternative window does not reveal a reliable holdout gain in QLIKE. With a fixed 30 day window, the ridge improvement is 0.0084 with a 95% interval of $[-0.0086, 0.0315]$, while the gradient boosting improvement is -0.0209 with an interval of $[-0.0740, 0.0262]$. With a fixed 100 day window, the corresponding improvements are 0.0020 with an interval of $[-0.0027, 0.0070]$ for ridge regression and 0.0003 with an interval of $[-0.0162, 0.0183]$ for gradient boosting. Every interval contains zero. The alternative losses likewise provide no interval excluding zero at either robustness window.

Complete fixed window robustness results are reported in Table S5 of the Supplementary Material.

The secondary procedure that permits validation to select the window reproduces the earlier holdout choices: 50 days for baseline ridge regression and 30 days for augmented ridge regression. It gives a QLIKE improvement of 0.010 with an interval that contains zero. This result is a valid comparison between two separately tuned forecasting procedures. It is not evidence that can be attributed solely to adding TDA, because changing the window also changes the classical rolling predictors. For this reason, it does not replace the fixed 50 day comparison as the confirmatory analysis.

Table S6 in the Supplementary Material reports the complete secondary pipeline comparison.

5.6 Assessment of incremental value

Regarding RQ4, the forecasting analysis does not establish a stable advantage across models, periods, and windows. Under the primary fixed 50 day design, ridge regression deteriorates in every development fold and in the final holdout. Its holdout QLIKE interval excludes zero in the unfavorable direction. The analyses with fixed windows of 30 and 100 days do not yield reliable QLIKE gains either. These results do not support a stable linear contribution after conventional variables have been included.

The nonlinear evidence is narrower. Gradient boosting improves average performance during development and all four holdout point estimates under the primary window, but only two of five development folds favor the augmented model. Its QLIKE interval includes zero, and the augmented model exceeds the naive forecast only under QLIKE. The favorable intervals for absolute and squared volatility error remain secondary results. Taken together, the evidence is consistent with conditional nonlinear value under selected losses, not a general forecasting advantage.

These results place the structural findings of Sections 3 and 4 in perspective. A variable may respond strongly to dependence across indices, retain variation that conventional indicators do not explain contemporaneously, and coincide with broad changes in market organization without improving a strict forward forecast. Incremental prediction imposes a more demanding requirement: the remaining

information must be stable, available at the forecast date, and usable by the selected learner. In this sample, the clearest role of persistent homology is therefore structural monitoring and regime characterization. Section 6 develops the interpretation of this distinction.

6 Discussion

6.1 Interpretation of the combined evidence

The results support a bounded interpretation of topological analysis in the Tehran Stock Exchange. The persistent homology measures respond to changes in the joint configuration of the five indices, differ across independently defined market regimes, and contribute to some forecasts produced by the nonlinear model. They do not, however, provide a uniformly superior forecasting signal. Statistical responsiveness, structural distinctiveness, and incremental forecasting value are separate propositions. The first receives strong support in this study. The surrogate experiments qualify the second, and the forecasting results show that the third depends on the model, period, and loss function. The evidence therefore supports the use of topology primarily as a compact representation of dependence geometry across the market rather than as an independent forecasting technology.

The concentration of the index panel helps explain this interpretation. Results based on the primary window show that larger effective correlation rank is generally associated with richer or stronger one dimensional topology. Conversely, greater concentration in the leading correlation component is associated with lower entropy, fewer loops, and weaker persistence measures. These relationships are consistent with the geometry of a standardized point cloud whose observations increasingly align with a common market direction. Persistent homology reorganizes information about dependence, dispersion, and effective dimension into a nonlinear geometric representation, but it does not necessarily supply information independent of those characteristics.

The surrogate analysis identifies the statistical source of this pattern more directly. Independent permutation preserves the marginal return series but removes their contemporaneous alignment, and this operation produces large differences from the observed topology. In contrast, surrogates based on Gaussian covariance and a Gaussian copula reproduce much more of the observed feature distributions. No comparison with either dependence preserving null remains significant after adjustment by the Benjamini and Hochberg procedure. The data therefore reject an explanation based on marginal behavior alone, but they do not establish that the observed topology contains a separate structure beyond that generated by the fitted Gaussian dependence models. This result identifies dependence as the main source of the topological signal and prevents a stronger claim that the persistence measures reveal an entirely distinct dimension of market information.

6.2 Relation to earlier financial TDA studies

Previous studies report that statistics based on persistence respond around episodes of financial turbulence and summarize changes in market organization (Gidea M, Katz Y , 2018; Aguilar A, Ensor KB , 2020; Baitinger E, Flegel S , 2021). Other work considers their use in detecting warning patterns, guiding investment decisions, or improving forecasts (Guo H, et al. , 2020; Ismail MS, et al. , 2022; de Jesus Jr LC, et al. , 2025). The present study complements this literature by separating three empirical questions. A topological measure may vary during financial stress, yet its variation may still

be reproducible under a conventional dependence model. Even when a residual component remains, it may not improve a forecast after classical predictors and model selection are taken into account. Visual responsiveness, statistical distinctiveness, and incremental prediction should therefore be evaluated separately.

The comparison of symbolic aggregate approximation and TDA by Hobbelhaugen F, Diamantis I (2024) is particularly relevant. The authors find that TDA describes local movements in greater detail, whereas their symbolic method captures broader market trends more effectively when less detail is required. The present results suggest a related dependence on the empirical task. The topological variables characterize changes in dependence geometry and differ across broad regimes, but this sensitivity does not translate consistently into lower forecast losses. TDA may therefore be useful for structural monitoring and nonlinear representation, while simpler temporal or econometric summaries can remain preferable when the objective is robust point prediction.

The Tehran market extends this evidence to a concentrated emerging market index system. The independently estimated regimes show that volatility and dependence need not move back toward their earlier levels at the same time. The boundary in June 2018 separates a low stress regime from a period with higher volatility, deeper drawdowns, stronger common movement, and lower effective rank. After the less stable boundary in April 2022, volatility and drawdown measures decline, but average correlation and concentration in the leading component increase further. This pattern represents partial normalization in the scale of fluctuations without comparable normalization in dependence. It also explains why the topological variables can remain structurally responsive when conventional measures of volatility moderate: the effective geometry of the market remains highly concentrated.

This interpretation is descriptive rather than causal. The change points were estimated retrospectively from classical indicators, and the associations between those indicators and the topological measures do not identify a transmission mechanism. The analysis shows that the two groups of variables evolve coherently with changes in market organization; it does not establish that one group causes the other or that either regime boundary would have been detected at the time.

6.3 Implications for risk measurement and forecasting

For market monitoring, the most defensible role of the topological measures is to complement volatility, correlation, concentration, and drawdown indicators. A monitoring system could display both groups of measures. Similar movements would provide evidence of a broad reorganization in market dependence. Divergence could identify periods that warrant closer examination, although the present study does not supply a formal decision rule for interpreting such divergence. Since agreement across windows of 30, 50, and 100 trading days is positive but moderate, the window should be treated as a temporal horizon. Shorter windows emphasize recent changes, whereas longer windows summarize more persistent organization and respond more slowly.

The forecasting evidence calls for similar restraint. Under the primary fixed 50 day design, adding the TDA variables raises ridge QLIKE in every development fold. It also raises holdout QLIKE, and the bootstrap interval lies below zero when the difference is expressed as an improvement. The gradient boosting model benefits in only two of the five development folds. In the untouched holdout it records lower losses under all four criteria, but its QLIKE interval includes zero. The intervals for improvements in absolute and squared volatility errors exclude zero. These are secondary results,

however, and the analysis does not adjust for examining several combinations of models and loss functions. Moreover, the gradient boosting model with TDA variables outperforms the naive forecast only under QLIKE; the naive forecast has lower errors under the other three reported criteria. The forecasting contribution is therefore conditional rather than general.

Fixing the information window is essential for attributing these differences. The confirmatory comparison uses 50 trading days for both variable sets, and the robustness analyses repeat the same restriction at 30 and 100 days. None of the alternative windows yields a reliable holdout QLIKE gain. A secondary analysis permits validation to choose the window separately for each variable set and selects 50 days for baseline ridge regression and 30 days for augmented ridge regression. That result compares complete forecasting procedures, but it cannot isolate the effect of TDA because the change in window also changes the classical rolling predictors. It is therefore not used as the confirmatory test.

The analysis also illustrates the importance of an untouched holdout and validation that respects temporal order. Overlapping windows create persistent predictors and dependent forecast errors. Searches across windows and model settings create opportunities to select favorable results, while relationships observed in one regime may weaken in another. An apparent gain from a separately selected window could also be misattributed to the added variables. Such results could easily be overstated without the fixed window comparison, the less favorable development evidence, the baseline models, and the paired block bootstrap. The relevant conclusion is not that topology succeeds or fails in general. Its practical value must be demonstrated for a defined target, information set, learner, and loss function.

6.4 Limitations

Several limitations determine the scope of the findings. First, the analysis covers one emerging equity market and only five aggregate indices. The strong common component is empirically important, but it also restricts the diversity of the point cloud. A panel of firms, industries, or markets may generate a richer geometry and a different balance between common and idiosyncratic variation. Results from this concentrated panel should not be assumed to apply to those settings.

Second, the representation uses daily returns, Euclidean distance, Vietoris Rips filtrations, standardization within each window, and one dimensional homology. Alternative embeddings, distance measures, filtrations, or homology dimensions may emphasize different structures. The present design is a transparent and predefined specification rather than an exhaustive search over topological constructions. This restraint reduces the risk of selecting a favorable representation, but it also means that the study cannot rank the full set of possible TDA methods.

Third, the surrogate experiments use a finite number of replications and a subsample of the overlapping windows. Adjustment for multiple comparisons controls the false discovery rate across the implemented tests, but it does not remove uncertainty due to finite simulation or serial dependence among adjacent windows. The finding that no comparison with the covariance or copula null survives adjustment should therefore be read as an absence of evidence under the implemented design, not as proof that the null models reproduce every aspect of the observed process.

Fourth, the regime analysis is retrospective. The boundaries are estimated without topological variables, which avoids circular validation, but standardization, segmentation, consensus formation,

and regime naming use the complete sample. The procedure cannot therefore be interpreted as a warning system that would have identified the June 2018 or April 2022 transitions on those dates. The first boundary is stable across all five segmentation designs, whereas recovery of the second boundary depends on the window and segmentation thresholds. This difference also prevents the two boundaries from receiving equal evidential weight.

Finally, the forecasting exercise considers one target, realized variance over the next 20 trading days, and two fitted algorithms. It does not establish performance for returns, tail losses, drawdowns, other horizons, or portfolio decisions after transaction costs. Although the last 530 observations form an untouched holdout, they represent only one recent period. Neither the absence of a universal gain nor the favorable secondary results for gradient boosting should be generalized beyond the evaluated sample, target, models, and criteria.

6.5 Directions for further research

The most informative extension would test whether the findings transfer to other panels. The same predefined pipeline could be applied to a broader set of Tehran industries and to external markets while retaining a separate holdout for every application. Such a design would show whether the dependence based interpretation is specific to a concentrated index panel or recurs in markets with greater diversity.

A second extension should convert the retrospective regime analysis into a prospective monitoring design. At each date, standardization, model selection, boundary estimation, and any alert threshold would use only information then available. Detection delay, false alerts, and stability after new observations arrive should be reported explicitly. This exercise would assess operational warning value; it would not, by itself, establish a causal interpretation.

A third extension could evaluate TDA for decisions in which dependence geometry has a direct role, including changes in covariance structure, loss of diversification, and classification of future tail risk. Each application should retain an incremental comparison: a classical benchmark and the same benchmark supplemented by a predefined set of topological variables. Evaluation should use temporally ordered validation, paired inference, adjustment where several secondary comparisons are examined, and losses tied to the intended decision. These requirements preserve the central lesson of this study: topological sensitivity can reveal systematic structural variation, but practical value must be established for a specific use rather than inferred from visual patterns or separation within the estimation sample.

7 Conclusions

This study examined whether persistent homology provides an informative representation of dependence structure and market regimes in the Tehran Stock Exchange and adds incremental predictive value for future realized variance. The empirical analysis used 2,769 daily observations for five major indices from 26 March 2014 to 22 September 2025. It combined topological measures computed over rolling windows, conventional indicators of volatility and dependence, surrogate experiments, regimes defined independently of topology, and forecasting tests that preserved temporal order and reserved a final holdout. This design separated four questions: RQ1, whether topological measures respond to changing market conditions; RQ2, to what extent their variation extends beyond conven-

tional descriptions of dependence; RQ3, how stable the identified market regimes are across alternative specifications; and RQ4, whether any remaining information improves forecasts.

The results provide qualified answers to RQ1 and RQ2. The persistent homology measures vary substantially through time and are associated with changes in market dependence and effective dimension. Their behavior does not arise from the marginal return series alone. Removing contemporaneous dependence through independent permutation produces large departures from the observed topology. However, surrogates based on Gaussian covariance and a Gaussian copula reproduce much more of the empirical feature distributions, and no comparison with either null remains significant after adjustment by the Benjamini and Hochberg procedure. The topological signal is therefore best understood as a compact nonlinear representation of dependence geometry, not as information wholly separate from correlation, concentration in the leading component, and effective rank.

Regarding RQ3, the regime analysis supports this structural interpretation without using the topological variables to define the regimes. A stable boundary in June 2018 separates the initial period of low stress from a period with higher volatility, deeper drawdowns, stronger common movement, and lower effective rank. The boundary in April 2022 receives less consistent support across the alternative designs. After that date, volatility and drawdown decline relative to the preceding regime, while average correlation and concentration in the leading component increase further. The market therefore exhibits partial normalization in the scale of fluctuations without a corresponding restoration of diversification. These regimes are retrospective descriptions and should not be interpreted as alerts that would necessarily have been available on the estimated boundary dates.

Regarding RQ4, the forecasting results offer more limited evidence. The confirmatory comparison fixes the window at 50 trading days for both variable sets. Adding the TDA variables raises ridge QLIKE in all five development folds and in the final holdout. The holdout bootstrap interval lies below zero when the difference is expressed as an improvement, indicating deterioration under the primary model and loss. Histogram gradient boosting performs better with TDA variables in only two of the five development folds but reduces all four holdout losses relative to the corresponding gradient boosting baseline. Its QLIKE interval nevertheless includes zero. The intervals for improvements in absolute and squared volatility errors exclude zero, but these are secondary comparisons and no adjustment was made for examining several combinations of models and losses. Fixed window analyses at 30 and 100 days also yield no reliable holdout QLIKE improvement.

The naive persistence forecast further limits the predictive claim. The gradient boosting model with TDA variables has a lower holdout QLIKE than the naive forecast, but the naive forecast has a lower mean absolute volatility error, a lower root mean squared volatility error, and a lower error for log variance. The augmented nonlinear model therefore does not dominate the naive benchmark across the reported criteria. The evidence supports conditional forecasting value for a flexible learner under selected losses, not a general forecasting advantage for TDA.

The study also contributes an evaluation framework for financial applications of topology. The framework compares topological measures with classical benchmarks, uses surrogate data to investigate their statistical origin, defines market regimes without reference to those measures, and tests incremental forecasting value with nested validation and a separate holdout. These steps reduce the risk of assigning economic or predictive importance to a measure only because it displays striking movements during turbulent periods.

For applied risk analysis, the persistent homology measures should complement established indicators rather than replace them. Their most defensible role in this sample is to summarize changes in the joint geometry of the market and to distinguish moderation in volatility from restoration of diversification. Any operational monitoring rule would require prospective validation using only information available at each date. Any forecasting application would likewise require evaluation for its specific target, horizon, model, benchmark, and loss function.

Overall, persistent homology is informative about the organization of the Tehran Stock Exchange, but its empirical value is more clearly structural than predictive. It provides a nonlinear view of dependence concentration and regime evolution, while its forecasting contribution remains sensitive to the learner and evaluation criterion. This bounded conclusion defines both the contribution of the method in the present study and the evidence still required before broader practical claims can be made.

Author contributions

Author 1 and Author 2: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing original draft, and Writing review and editing. Both authors read and approved the final manuscript.

Use of AI tools declaration

OpenAI ChatGPT was used to assist with language editing and LaTeX preparation. The authors verified all methodological choices, computations, numerical results, references, and interpretations and retain full responsibility for the manuscript. No empirical observation or numerical result was generated by the AI tool.

Acknowledgments

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The replication archive supplied with the manuscript contains the analysis code, the return panel used in the analysis, the TEDPIX series used for source validation, intermediate results, and the files required to reproduce the reported tables and figures. Historical index levels were retrieved through the FinPy-TSE Python package. Users retrieving the source data independently remain subject to the

terms of the underlying TSE data services accessed by that package. Public release of the replication materials will follow the journal’s data and code availability policy.

References

- Aguilar A, Ensor KB (2020) Topology data analysis using mean persistence landscapes in financial crashes. *J Math Financ* 10: 648–678. <https://doi.org/10.4236/jmf.2020.104038>
- Akingbade SW, Gidea M, Manzi M, Nateghi V (2024) Why topological data analysis detects financial bubbles? *Commun Nonlinear Sci Numer Simul* 128: 107665. <https://doi.org/10.1016/j.cnsns.2023.107665>
- Baitinger E, Flegel S (2021) The better turbulence index? Forecasting adverse financial markets regimes with persistent homology. *Financ Mark Portf Manag* 35: 277–308. <https://doi.org/10.1007/s11408-020-00377-x>
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57: 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bubenik P (2015) Statistical topological data analysis using persistence landscapes. *J Mach Learn Res* 16: 77–102.
- Carlsson G (2009) Topology and data. *Bull Am Math Soc* 46: 255–308. <https://doi.org/10.1090/S0273-0979-09-01249-X>
- Corsi F (2009) A simple approximate long-memory model of realized volatility. *J Financ Econom* 7: 174–196. <https://doi.org/10.1093/jjfinec/nbp001>
- de Jesus Jr LC, Fernández-Navarro F, Carbonero-Ruz M (2025) Enhancing financial time series forecasting through topological data analysis. *Neural Comput Appl* 37: 6527–6545. <https://doi.org/10.1007/s00521-024-10787-x>
- Edelsbrunner H, Harer J (2010) *Computational Topology: An Introduction*. Providence, RI: American Mathematical Society.
- Rahimi A, Ghafouri R (2024) FinPy-TSE: A Python package to access Tehran Stock Exchange historical and real-time data, version 1.2.10. Available from: <https://pypi.org/project/finpy-tse/> (accessed 7 September 2026).
- Gidea M, Katz Y (2018) Topological data analysis of financial time series: Landscapes of crashes. *Physica A* 491: 820–834. <https://doi.org/10.1016/j.physa.2017.09.028>
- Guo H, Xia S, An Q, Zhang X, Sun W, Zhao X (2020) Empirical study of financial crises based on topological data analysis. *Physica A* 558: 124956. <https://doi.org/10.1016/j.physa.2020.124956>
- Hobbelhagen F, Diamantis I (2024) A comparative study of symbolic aggregate approximation and topological data analysis. *Quant Financ Econ* 8: 705–732. <https://doi.org/10.3934/QFE.2024027>

- Ismail MS, Md Noorani MS, Ismail M, Abdul Razak F, Alias MA (2022) Early warning signals of financial crises using persistent homology. *Physica A* 586: 126459. <https://doi.org/10.1016/j.physa.2021.126459>
- Killick R, Fearnhead P, Eckley IA (2012) Optimal detection of changepoints with a linear computational cost. *J Am Stat Assoc* 107: 1590–1598. <https://doi.org/10.1080/01621459.2012.737745>
- Künsch HR (1989) The jackknife and the bootstrap for general stationary observations. *Ann Stat* 17: 1217–1241. <https://doi.org/10.1214/aos/1176347265>
- Newey WK, West KD (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55: 703–708. <https://doi.org/10.2307/1913610>
- Patton AJ (2011) Volatility forecast comparison using imperfect volatility proxies. *J Econometrics* 160: 246–256. <https://doi.org/10.1016/j.jeconom.2010.03.034>
- Perea JA, Harer J (2015) Sliding windows and persistence: An application of topological methods to signal analysis. *Found Comput Math* 15: 799–838. <https://doi.org/10.1007/s10208-014-9206-z>
- Valdivia AD (2023) Topological variability in financial markets. *Quant Financ Econ* 7: 391–402. <https://doi.org/10.3934/QFE.2023019>

A Supplementary tables

Table S1: Descriptive audit of the four prespecified TDA features.

Window	Feature		n	Mean	SD	Median	Zero share	Lag ACF	1 Lag ACF	20
30	Landscape (H_1)	L_2 norm	2740	0.029	0.029	0.019	0.014	0.876	0.128	
30	Maximum persistence (H_1)	persis-	2740	0.177	0.127	0.149	0.014	0.871	0.144	
30	Number of features	of H_1 fea-	2740	3.829	1.942	4.000	0.014	0.870	0.324	
30	Persistent (H_1)	entropy	2740	0.969	0.512	1.029	0.103	0.856	0.293	
50	Landscape (H_1)	L_2 norm	2720	0.043	0.030	0.035	0.000	0.928	0.364	
50	Maximum persistence (H_1)	persis-	2720	0.226	0.114	0.203	0.000	0.910	0.311	
50	Number of features	of H_1 fea-	2720	8.762	2.985	8.000	0.000	0.927	0.392	
50	Persistent (H_1)	entropy	2720	1.756	0.394	1.796	0.003	0.927	0.355	
100	Landscape (H_1)	L_2 norm	2670	0.067	0.039	0.059	0.000	0.968	0.520	
100	Maximum persistence (H_1)	persis-	2670	0.284	0.130	0.264	0.000	0.957	0.476	
100	Number of features	of H_1 fea-	2670	24.213	4.625	24.000	0.000	0.962	0.635	
100	Persistent (H_1)	entropy	2670	2.767	0.229	2.787	0.000	0.966	0.587	

Table S2: Surrogate-test summary. Fractions are proportions of evaluated windows.

Window	Surrogate	Feature	n	Observed	Surrogate mean	Mean z	Outside 95%	BH significant
30	Gaussian cop-ula	Landscape norm (H_1)	L_2 138	0.031	0.043	-0.428	0.101	0.000
30	Gaussian cop-ula	Maximum persistence (H_1)	138	0.185	0.231	-0.413	0.109	0.000
30	Gaussian cop-ula	Number of H_1 features	138	3.725	4.633	-0.529	0.014	0.000
30	Gaussian cop-ula	Persistent entropy (H_1)	138	0.934	1.151	-0.521	0.029	0.000
30	Gaussian variance	Landscape norm (H_1)	L_2 138	0.031	0.040	-0.342	0.043	0.000
30	Gaussian variance	Maximum persistence (H_1)	138	0.185	0.220	-0.325	0.036	0.000
30	Gaussian variance	Number of H_1 features	138	3.725	4.257	-0.309	0.007	0.000
30	Gaussian variance	Persistent entropy (H_1)	138	0.934	1.065	-0.307	0.043	0.000
30	Marginal independence	Landscape norm (H_1)	L_2 138	0.031	0.091	-1.365	0.500	0.283
30	Marginal independence	Maximum persistence (H_1)	138	0.185	0.367	-1.427	0.457	0.239
30	Marginal independence	Number of H_1 features	138	3.725	7.831	-1.792	0.399	0.225
30	Marginal independence	Persistent entropy (H_1)	138	0.934	1.696	-2.520	0.478	0.254
50	Gaussian cop-ula	Landscape norm (H_1)	L_2 545	0.043	0.062	-0.705	0.182	0.000
50	Gaussian cop-ula	Maximum persistence (H_1)	545	0.225	0.278	-0.566	0.167	0.000
50	Gaussian cop-ula	Number of H_1 features	545	8.758	10.932	-0.833	0.105	0.000
50	Gaussian cop-ula	Persistent entropy (H_1)	545	1.757	2.013	-0.946	0.161	0.000
50	Gaussian variance	Landscape norm (H_1)	L_2 545	0.043	0.059	-0.600	0.141	0.000
50	Gaussian variance	Maximum persistence (H_1)	545	0.225	0.271	-0.489	0.139	0.000
50	Gaussian variance	Number of H_1 features	545	8.758	10.229	-0.581	0.075	0.000
50	Gaussian variance	Persistent entropy (H_1)	545	1.757	1.940	-0.665	0.125	0.000
50	Marginal independence	Landscape norm (H_1)	L_2 545	0.043	0.122	-1.957	0.661	0.611
50	Marginal independence	Maximum persistence (H_1)	545	0.225	0.408	-1.694	0.560	0.457
50	Marginal independence	Number of H_1 features	545	8.758	16.572	-2.440	0.635	0.497
50	Marginal independence	Persistent entropy (H_1)	545	1.757	2.451	-3.418	0.661	0.532
100	Gaussian cop-ula	Landscape norm (H_1)	L_2 135	0.069	0.088	-0.859	0.378	0.000
100	Gaussian cop-ula	Maximum persistence (H_1)	135	0.290	0.305	-0.204	0.215	0.000
100	Gaussian cop-ula	Number of H_1 features	135	24.133	31.086	-1.648	0.370	0.000
100	Gaussian cop-ula	Persistent entropy (H_1)	135	2.761	3.070	-2.014	0.481	0.000

Window	Surrogate	Feature	n	Observed	Surrogate mean	Mean z	Outside 95%	BH significant
100	Gaussian variance	co- Landscape norm (H_1)	L_2 135	0.069	0.084	-0.717	0.311	0.000
100	Gaussian variance	co- Maximum persistence (H_1)	135	0.290	0.301	-0.154	0.200	0.000
100	Gaussian variance	co- Number of H_1 features	135	24.133	29.414	-1.286	0.237	0.000
100	Gaussian variance	co- Persistent entropy (H_1)	135	2.761	3.009	-1.575	0.356	0.000
100	Marginal independence	inde- Landscape norm (H_1)	L_2 135	0.069	0.175	-2.902	0.830	0.807
100	Marginal independence	inde- Maximum persistence (H_1)	135	0.290	0.450	-1.713	0.600	0.504
100	Marginal independence	inde- Number of H_1 features	135	24.133	43.231	-3.870	0.844	0.830
100	Marginal independence	inde- Persistent entropy (H_1)	135	2.761	3.413	-5.406	0.881	0.867

Table S3: Primary incremental comparisons from the fixed window of 50 trading days.

Model	Metric	Folds	Mean gain	Median gain	Positive-fold share
HGB	QLIKE	5	0.0465	-0.0072	0.40
HGB	MAE volatility	5	0.0009	-0.0013	0.40
HGB	RMSE volatility	5	0.0001	-0.0025	0.40
HGB	RMSE log variance	5	0.0105	-0.0096	0.40
ridge	QLIKE	5	-0.0337	-0.0291	0.00
ridge	MAE volatility	5	-0.0004	-0.0006	0.40
ridge	RMSE volatility	5	-0.0005	-0.0013	0.40
ridge	RMSE log variance	5	-0.0138	-0.0144	0.20

Table S4: Metrics for individual folds under the primary forecasting design with a fixed window of 50 trading days.

Fold	Model	Variables	Window Parameters		QLIKE	MAE vol.	RMSE vol.
1	ridge	Baseline	50	$\alpha = 100$	-4.3723	0.0345	0.0372
1	ridge	Baseline + TDA	50	$\alpha = 100$	-4.2978	0.0359	0.0391
1	HGB	Baseline	50	leaves=7; $L_2 = 10$	-3.8863	0.0680	0.0766
1	HGB	Baseline + TDA	50	leaves=15; $L_2 = 1$	-3.8751	0.0697	0.0798
2	ridge	Baseline	50	$\alpha = 100$	-0.5158	0.0963	0.1339
2	ridge	Baseline + TDA	50	$\alpha = 100$	-0.4792	0.0953	0.1320
2	HGB	Baseline	50	leaves=7; $L_2 = 10$	2.4700	0.1174	0.1477
2	HGB	Baseline + TDA	50	leaves=7; $L_2 = 1$	2.2907	0.1142	0.1445
3	ridge	Baseline	50	$\alpha = 100$	-1.0258	0.0960	0.1263
3	ridge	Baseline + TDA	50	$\alpha = 100$	-0.9967	0.0977	0.1288
3	HGB	Baseline	50	leaves=7; $L_2 = 1$	-0.4668	0.1277	0.1484
3	HGB	Baseline + TDA	50	leaves=7; $L_2 = 1$	-0.5923	0.1206	0.1425
4	ridge	Baseline	50	$\alpha = 100$	-1.7764	0.0633	0.0805
4	ridge	Baseline + TDA	50	$\alpha = 100$	-1.7754	0.0626	0.0793
4	HGB	Baseline	50	leaves=7; $L_2 = 10$	-1.9014	0.0489	0.0617
4	HGB	Baseline + TDA	50	leaves=7; $L_2 = 10$	-1.8942	0.0516	0.0643
5	ridge	Baseline	50	$\alpha = 100$	-2.1476	0.0554	0.0757
5	ridge	Baseline + TDA	50	$\alpha = 100$	-2.1205	0.0560	0.0769
5	HGB	Baseline	50	leaves=7; $L_2 = 10$	-1.9736	0.0574	0.0754
5	HGB	Baseline + TDA	50	leaves=7; $L_2 = 1$	-1.9198	0.0587	0.0783

Table S5: Holdout incremental value under robustness designs using fixed windows of 30 and 100 trading days.

Window	Model	Metric	Mean gain	95% interval	$P(\text{gain} > 0)$
30	ridge	QLIKE	0.00844	[-0.00861, 0.03146]	0.789
30	ridge	Absolute volatility error	-0.00032	[-0.00152, 0.00115]	0.298
30	ridge	Squared volatility error	-0.00000	[-0.00013, 0.00015]	0.477
30	HGB	QLIKE	-0.02087	[-0.07402, 0.02622]	0.201
30	HGB	Absolute volatility error	0.00116	[-0.00315, 0.00540]	0.696
30	HGB	Squared volatility error	0.00020	[-0.00026, 0.00066]	0.799
100	ridge	QLIKE	0.00200	[-0.00273, 0.00696]	0.773
100	ridge	Absolute volatility error	0.00021	[-0.00042, 0.00083]	0.739
100	ridge	Squared volatility error	0.00003	[-0.00003, 0.00008]	0.855
100	HGB	QLIKE	0.00030	[-0.01624, 0.01828]	0.489
100	HGB	Absolute volatility error	0.00130	[-0.00212, 0.00487]	0.757
100	HGB	Squared volatility error	0.00021	[-0.00021, 0.00068]	0.827

Table S6: Secondary comparison of separately tuned forecasting pipelines. This table does not isolate the contribution of TDA because the selected windows may differ.

Model	Metric	Baseline win- dow	TDA window	Mean gain	95% interval
ridge	QLIKE	50	30	0.01007	[-0.02708, 0.05962]
ridge	Absolute volatility error	50	30	-0.00070	[-0.00331, 0.00212]
ridge	Squared volatility error	50	30	-0.00003	[-0.00034, 0.00031]
HGB	QLIKE	50	50	0.02701	[-0.00862, 0.06253]
HGB	Absolute volatility error	50	50	0.00463	[0.00038, 0.00930]
HGB	Squared volatility error	50	50	0.00055	[0.00011, 0.00106]

Table S7: Stability of the eight audited topological features across window lengths.

Window A	Window B	Common dates	Median $ r $	Minimum $ r $	Median $ \rho $	Minimum $ \rho $	Share $ \rho \geq 0.50$
30	50	2720	0.450	0.263	0.496	0.331	0.500
30	100	2670	0.294	0.202	0.356	0.230	0.125
50	100	2670	0.433	0.264	0.468	0.307	0.375