The Journal of Finance and Data Science 9 (2023) 100107

Contents lists available at ScienceDirect

The Journal of Finance and Data Science

journal homepage: www.keaipublishing.com/en/journals/jfds

Research Article

Topological tail dependence: Evidence from forecasting realized volatility

Hugo Gobato Souto

International School of Business at HAN University of Applied Sciences, Ruitenberglaan 31, 6826 CC Arnhem, the Netherlands

ARTICLE INFO

JEL classification:
Neural networks (C45)
NBEATSx (C45)
Realized volatility forecasting (C53)

Keywords:
Neural networks
NBEATSx
Realized volatility forecasting

ABSTRACT

This paper proposes a novel theory, coined as Topological Tail Dependence Theory, that links the mathematical theory behind Persistent Homology (PH) and the financial stock market theory. This study also proposes a novel algorithm to measure topological stock market changes as well as the incorporation of these topological changes into forecasting realized volatility (RV) models to improve their forecast performance during turbulent periods. The results of the empirical experimentation of this study provide evidence that the predictions drawn from the Topological Tail Dependence Theory are correct and indicate that the employment of PH information allows nonlinear and neural network models to better forecast RV during a turbulent period.

1. Introduction

Topological data analysis (TDA) encompasses a growing set of methods that explore the topology, also commonly referred to as geometry, of data and leverage this topological information to infer patterns from the data. Since its creation, the field of TDA has demonstrated considerable potential in a wide variety of domains, particularly in those that handle high-dimensional data (Bubenik, 2015; Carlsson, 2009; Wasserman, 2017). TDA's success ranges from medical research (McInemey and Terzopoulos, 1999; Nicolau et al., 2011) to financial research (Baitinger and Flegel, 2021; Gidea and Katz, 2018; Ismail et al., 2022). It is noteworthy that the majority of TDA literature utilizes the technique known as Persistent Homology (PH). PH is a remarkably effective algorithm that attempts to quantitatively find elementary persistent topological structures, patterns and qualitative features in data that are independent of coordinate representations and deformations related to noise (Baitinger and Flegel, 2021). By employing PH, researchers can uncover hidden yet fundamental patterns in data, leading to novel discoveries about patterns in the data (Bubenik, 2015; Carlsson, 2009; Wasserman, 2017).

While numerous fields have extensively researched the applications of TDA since the beginning of the 21st century, the finance literature had its first TDA research paper only in 2017 (Gidea, 2017). However, since then, a substantial amount of financial literature on TDA has been produced focused on topics ranging from corporate and systemic financial stability (Gidea, 2017; Gidea and Katz, 2018; Ismail et al., 2022; Qiu et al., 2020) to investment strategies (Baitinger and Flegel, 2021; Goel et al., 2020).

Nonetheless, despite the major importance of modeling and forecasting stock realized volatility (RV) for investors and financial institutions (Albulescu, 2021; Atkins et al., 2018; Jebran et al., 2017; Wong et al., 2016), there currently exists no financial literature exploring TDA's possible applications in this area. Additionally, there is currently no theory that creates a link between the theoretical background of PH and finance theory. Although empirically the implementation of PH to foresee financial distressful periods and to

E-mail address: hugo.gobatosouto@han.nl.

Peer review under responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.jfds.2023.100107
Received 13 June 2023; Received in revised form 15 September 2023; Accepted 6 October 2023
Available online 14 October 2023

improve portfolio management has already been proven successful, the same still needs to be done theoretically. Therefore, this paper has the objective to partially close both gaps in the scientific community by proposing the employment of PH in forecasting RV, proposing a novel theory that links the theoretical background of PH and finance theory, and testing the consequences of the proposed theory through an extensive empirical experimentation. The main contributions of this paper to the scientific community include:

(i) **Proposition of a novel algorithm to measure topological changes of stock time-series:** this research modifies the topological change metric PH-based turbulence index (PHTI), originally proposed by Baitinger and Flegel (2021), to adapt it to be effectively employed in the prediction of RV.
(ii) **Proposition of a novel and superior forecasting model:** this paper incorporates the proposed topological change metric into linear, nonlinear, and neural network models for forecasting realized volatility. This study demonstrates that the proposed incorporation achieves statistically significantly more accurate forecasts than the existing state-of-the-art models in financially turbulent periods for all models.
(iii) **Provision of open-source code for the proposed algorithm and forecasting model:** this study supplies open-source code for an easy implementation of the proposed algorithm and forecasting model by researchers and practitioners.
(iv) **Proposition of a novel theory that links the theoretical background of PH and finance theory:** this research proposes a novel theory coined "Topological Tail Dependence Theory" to explain the reason behind the recent success of the implementation of PH to foresee financial distressful periods and to improve portfolio management. Additionally, this theory is put to the test through an extensive empirical experimentation.

2. Persistent Homology and proposed algorithm

2.1. Persistent Homology

To ensure sparsity, this paper will solely provide a brief yet intuitive explanation without much mathematical rigor about PH and the proposed algorithm to measure topological chances of stock time-series for RV forecasting. Therefore, for a technical and mathematically rigorous explanation about PH, please see Edelsbrunner and Harer (2009).

2.1.1. Betti numbers

Before introducing PH, Betti numbers must be explained. The dth Betti number, β_d, is essentially the number of d-dimensional holes in a certain topological shape, also called simplicial complex (SC). For instance, β_0 can be thought of as the number of connected components (geometric elements and points), whereas β_1 as the number of (one-dimensional) holes and β_2 as the number of 3-d spaces that are encapsulated (voids). The elements of a SC, on the other hand, are called simplices, and all SCs can be considered a mathematical group. For instance, the SC in Fig. 1 can be written as K = {{a}, {b}, {c}, {a, b}, {b, c}, {a, c}}.

Graphically, it is obvious that $\beta_0 = 1$ (one triangle), $\beta_1 = 1$ (one one-dimensional hole), and $\beta_2 = 0$ (zero voids). Interestingly, the SC in Fig. 2 also has the exact same Betti numbers. This is the case since unlike in geometry, in topology we are concerned with the number of holes in different dimensions and not angles and lengths.

However, as the number of dimensions of a SC increases above three dimensions, it is no longer possible to visually determine its Betti numbers. Thus, one can use the following equations to mathematically estimate β_d:

$$\beta_d = rank\ H_d \tag{1}$$

$$H_d = \ker \partial_d / \mathrm{im}\partial_{d+1} \tag{2}$$

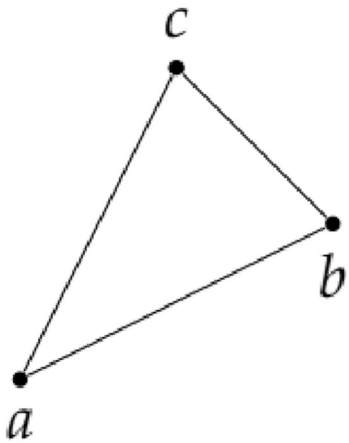

Fig. 1. Triangle.

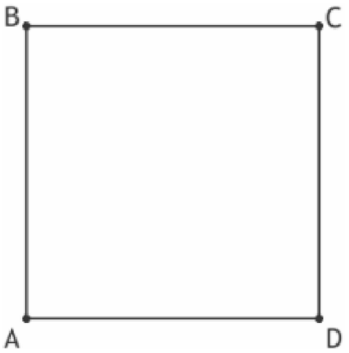

Fig. 2. Square.

$$\partial_d(K) = \sum_m \{v_0, \ldots, \widehat{v_m}, \ldots, v_d\} \tag{3}$$

where, ∂_d is the dth boundary homomorphism function applied to K, ker is the kernel of a homomorphism and im is the image of a homomorphism. Although these formulas can seem quite complex for someone without advanced knowledge in abstract algebra and linear algebra, their true meaning is simple. To visualize their true meaning, let us apply them to K. Recall that K = {{a}, {b}, {c}, {a, b}, {b, c}, {a, c}}. Starting with ∂_d, if we consider $\partial_0(K)$, we should consider only the zero-dimensional subsets (i.e., subsets with the points) of K, namely {{a}, {b}, {c}}. Essentially, ∂_0 is saying that for each subset we should create a sum of the elements inside the subset with each sum term being equal to the sum of all elements besides a certain element. Since each subset of {{a}, {b}, {c}} has only one element, all sums will simply be equal to zero.

$$\partial_0(K) = \{0, 0, 0\}$$

Now considering $\partial_1(K)$, we should only consider the one-dimensional subsets {{a, b}, {b, c}, {a, c}}. By creating a sum of the elements inside each subset with each sum term being equal to the sum of all elements besides a certain element, we have:

$$\partial_1(K) = \{\{a\} + \{b\}, \{b\} + \{c\}, \{a\} + \{c\}\}$$

ker ∂_d is essentially all terms from the originally considered subset that were dropped; so, $\ker\partial_0(K)$ = {{a}, {b}, {c}}. $\mathrm{im}\partial_{d+1}$, on the other hand, is all terms from the original subset that were in the final equation; thus, $\mathrm{im}\partial_1(K) = \{\{a\} + \{b\}, \{b\} + \{c\}, \{a\} + \{c\}\}$. Incidentally, $\ker\partial_0(K)$ and $\mathrm{im}\partial_1(K)$ can also be written as:

$$\ker\partial_0(K) = \mathrm{span}(\{a\}, \{b\}, \{c\}) = (\mathbb{Z}/2\mathbb{Z})^3$$

$$\mathrm{im}\partial_1(K) = \mathrm{span}(\{a\} + \{b\}, \{b\} + \{c\}) = (\mathbb{Z}/2\mathbb{Z})^2$$

where, $\mathbb{Z}/2\mathbb{Z}$ is the cyclic group of order 2 and $\mathrm{span}(\{a\}+\{b\}, \{b\}+\{c\}) = \mathrm{span}(\{a\}+\{b\}, \{b\}+\{c\}, \{a\}+\{c\})$ since $\{a\}+\{c\} = \{a\}+\{b\}+\{b\}+\{c\}$ in $\mathbb{Z}/2\mathbb{Z}$. To be able to explain the notations and concepts above, some prior knowledge about linear algebra and group theory would be needed. However, for the sake of simplicity, let us ignore the mathematical rigor enclosed in the equations above, and imagine the following: n in $(\mathbb{Z}/2\mathbb{Z})^n$ is simply determined by the number of elements in the span of $\ker\partial_0(K)$ and $\mathrm{im}\partial_1(K)$, and the span is the simplest way that one can write $\ker\partial_0(K)$ and $\mathrm{im}\partial_1(K)$ under the assumption that you can sum the elements in the span and $\{a\}+\{a\} = \{b\}+\{b\} = \{c\}+\{c\} = 0$. In other words, the sum of two equal elements must equal zero; thus, one can reconstruct $\{\{a\}+\{b\}, \{b\}+\{c\}, \{a\}+\{c\}\}$ by only using $\mathrm{span}(\{a\}+\{b\}, \{b\}+\{c\})$ as $\{a\}+\{c\} = \{a\}+\{b\}+\{b\}+\{c\}$.

Now regarding H_d, we can think of it as the ratio $\ker\partial_d/\mathrm{im}\partial_{d+1} = (\mathbb{Z}/2\mathbb{Z})^n/(Z/2Z)^m = (\mathbb{Z}/2\mathbb{Z})^{n-m}$. Hence, in our example of K, we have:

$$H_0(K) = (Z/2Z)^3/(\mathbb{Z}/2\mathbb{Z})^2 = \mathbb{Z}/2\mathbb{Z}$$

Lastly, *rank* H_d can be thought of as being equal to $n - m$. Consequently, in our example, we have:

$$\beta_0(K) = rank\ H_0(K) = 1$$

Although now you know how to analytically perform the estimation of Betti numbers given a certain SC, this procedure can be difficult to do by hand. Fortunately, there are currently various user-friendly programming packages for Python, R, C, etc. that democratize the estimation of Betti numbers for people without much knowledge about topology (Moroni and Pascali, 2021).

2.1.2. Persistent Homology

When dealing with real data, the correct SC is a priori not known; one only has single datapoints and does not possess their connections. As a result, there exists a great number of possible SCs, possibly each with different Betti numbers. Needless to say, this creates a problem when trying to analyze the topological structures of the data. To solve this issue, the PH technique and its graphs are utilized. To understand how PH graphs are used for this issue, consider the data points $D = \{d_1, \dots, d_n\}$ of an arbitrary circular and noiseless dataset and a distance metric *dist*. By picking a threshold ε, one can build the so-called Vietoris-Rips complex $\mathscr{V}_\varepsilon$ defined as:

$$\mathscr{V}_\varepsilon(D) := \{\sigma \subseteq D | \forall u, v \in \sigma : dist(u, v) \leq \varepsilon\} \tag{4}$$

In other words, $\mathscr{V}_\varepsilon$ contains all simplices (geometric elements such as {a}, {a, b}, {a, b, c}, etc.) whose diameter is less than or equal to ε. In simpler terms, we first plot the data points, then we choose a positive real value for ε, which is used to create a disk (or sphere, depending on the number of dimensions) around the data points with diameter equal to ε. If the disks (spheres) of two data points touch each other (i.e., the distance between these two data points is less than ε) then we create a link between them. In Fig. 3, Vietoris-Rips complexes for different ε's can be found for the data points D.

When estimating Vietoris-Rips complexes for N thresholds, as summarized in Fig. 4, Betti numbers can be estimated for all Vietoris-Rips complexes.

As a result, Fig. 5 can be constructed by estimating $\beta_1(\mathscr{V}_{\varepsilon_i}) \forall i \in \{1, \dots, N\}$ for the Vietoris-Rips complexes of Fig. 4.

It can be seen that until $\varepsilon_i = 0.2$, $\beta_1 = 0$, then a β_1 is "created" and lasts until $\varepsilon_i = 1$, where it is "destructed". The concepts of creation and destruction of Betti numbers are crucial for the construction of PH graphs. Creation in ε_i can be mathematically defined as $c \in \beta_d(\mathscr{V}_{\varepsilon_i})$, but $c \notin \beta_d(\mathscr{V}_{\varepsilon_{i-1}})$, and destruction in ε_i as $c \notin \beta_d(\mathscr{V}_{\varepsilon_i})$, but $c \in \beta_d(\mathscr{V}_{\varepsilon_{i-1}})$, where $c := \{0, 1, \dots, C\}$ and $\beta_d(\mathscr{V}_{\varepsilon_i}) := \{0, 1, \dots, \beta_d\}$. Hence, in Fig. 4, $c = \{0, 1\}$, and there is a creation in $\varepsilon_i = 0.2$ and a destruction in $\varepsilon_i = 1$. Finally, by accounting for the creation and destruction thresholds for each c, the PH graph depicted in Fig. 6 can be created.

A real-life example of a PD graph that is used in this research is depicted in Fig. 7. Fig. 7 shows the PD of the log returns of the basket of stocks composed of Standard & Poor's 500 (S&P 500), Dow Jones Industrial Average (DJIA) and Russell 2000 (RUT) from the 1st of February 2000 until 21 March 2000. The H_0 points can be ignored since they represent single data points that, needless to say, do not show any important topological structure. Yet, while H_1 points close to the diagonal are simply topological structures created by noise (Bubenik, 2015; Carlsson, 2009; Wasserman, 2017), the H_1 points far from the diagonal can show that the data has a (or some) topological structure(s) that ought to be exploited.

2.1.3. Wasserstein distance

The last mathematical concept to be presented is the Wasserstein distance (WD) defined by Equation (5). This metric is used to estimate how different two PD graphs are from each other, and thus calculate the topological difference between two datasets or time frames.

$$W_d(PD_1, PD_2) := \left(\underbrace{\inf}_{\eta: PD_1 \to PD_2} \sum_{x \in PD_1} \|x - \eta(x)\|_\infty^d \right)^{\frac{1}{d}} \tag{5}$$

where, inf is the infimum operator, η is an arbitrary bijection (function) mapping elements of PD_1 onto PD_2, and d is the considered dimension. Unfortunately, to fully understand Equation (5), some knowledge of set theory would be needed. Yet, in simple terms, the WD between two PH graphs can be defined as how different two PH graphs are from each other (i.e., how topologically different two datasets are from each other). Fortunately, there are currently various user-friendly programming packages for Python, R, C, etc. that democratize its estimation for people without much knowledge about set theory (Moroni and Pascali, 2021).

In simple terms, PH can be summarized as the procedure of finding topological features of data by 1. plotting its datapoints, 2. surrounding the datapoints by discs or spheres of diameter ε and connecting datapoints if the respective datapoints circles touch each other, 3. estimating the number of holes for d dimensions with Betti numbers, 4. repeating this process for various values of ε, and finally 5. creating a PD graph with the creation and destruction ε values for each Betti number in each dimension. Afterwards, it is possible to compare the topological features of two datasets or time frames by comparing their respective PD graphs with each other through WD.

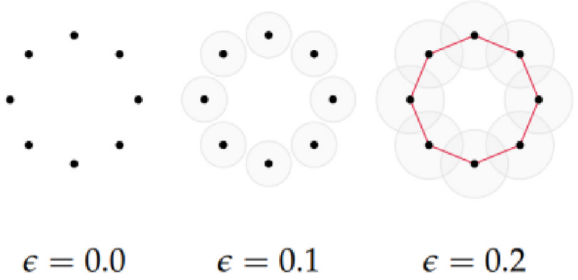

Fig. 3. Vietoris-Rips complexes for different thresholds.

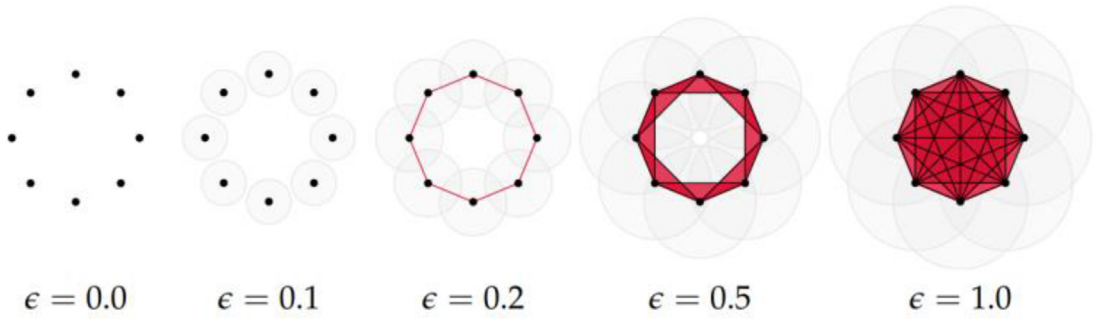

Fig. 4. Vietoris-Rips complexes for five different thresholds.

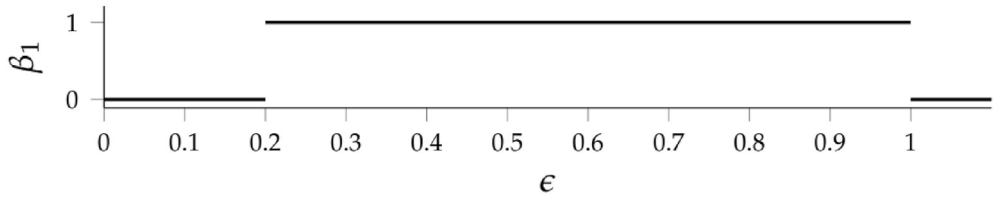

Fig. 5. 1st Betti number for each threshold.

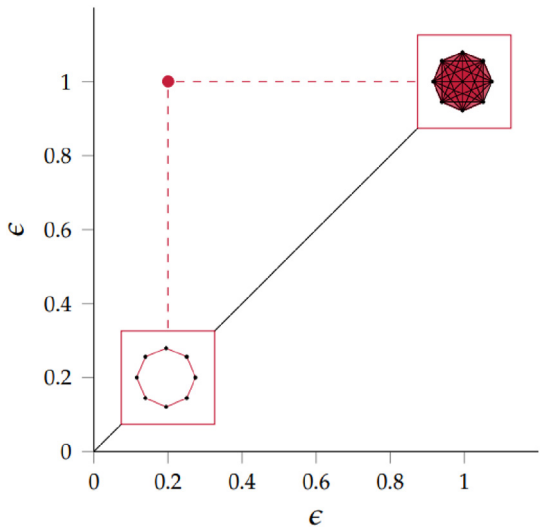

Fig. 6. Persistence homology graph for β_1.

2.2. Proposed algorithm

The proposed algorithm, coined as Persistent Homology for Realized Volatility (PH-RV), can be summarized as:

1. Construct a sparse basket of stocks containing stocks that are logically similar to each other, e.g. same industry, same economic region, close competitors, etc.
2. For the basket of stocks $S_t = \{s_t^1, \dots, s_t^n\}$ ranging from t_0 to T, estimate the log-return of each stock $R_t = \{r_t^1, \dots, r_t^n\}$ at each time step and m, which is equal to $T - 41$.
3. Estimate m numbers of PD graphs, each limited to a maximum two dimensions for the Betti numbers, by using the dataset $= \{R_{t-20}, \dots, R_t\}$ for each considered time step. That is, estimate PD graphs for m days using last month's log returns of each stock in the stock basket.
4. For m days, estimate the 2-dimensional Wasserstein distance between PD_t and PD_{t-1}.
5. Utilize the 2-dimensional Wasserstein distances as an exogenous variable for the chosen RV forecasting model.

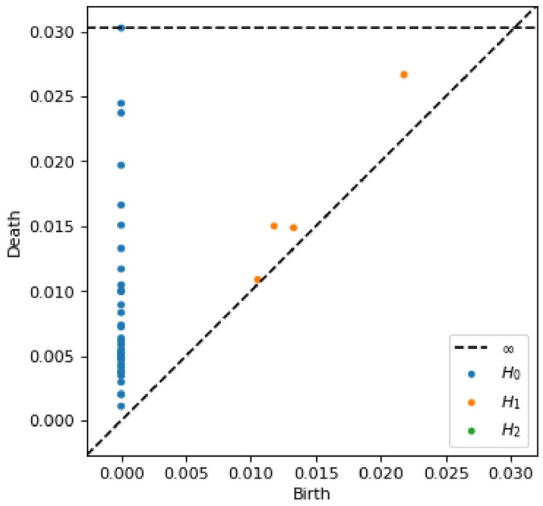

Fig. 7. PD of S&P 500, DJIA, and RUT log returns.

It is worth mentioning that since the ensemble of stock log-returns can be viewed as a point cloud, the features needed for the utilization of PH with such data are met.

The Python open-source code provided by this research for this algorithm can be found on https://github.com/hugogobato/Topological-Tail-Dependence-Evidence-from-Forecasting-Realized-Volatility/blob/4e4ae5fcdeb692228c009d70d1316bd9e24dced3/Persistent%20Homology%20for%20Realized%20Volatility%20(PH-RV)%20algorithm.py. This open-source code employs the library devised by Tralie et al. (2018) and allows the utilization of Yahoo Finance's API for fast extraction of stock data.

3. Topological Tail Dependence Theory

Interestingly, there is currently no financial theory linking the mathematical theory behind PH and the financial stock market, despite the recent promising development in research on application of PH through WD or L^n norms of Persistent Landscape (for more information, see Gidea and Katz (2018)) into foreseeing financially turbulent periods (Gidea, 2017; Gidea and Katz, 2018; Ismail et al., 2022; Qiu et al., 2020) and into portfolio management (Baitinger and Flegel, 2021; Goel et al., 2020). Consequently, this paper proposes a financial theory coined "Topological Tail Dependence Theory" in an attempt to close this existing scientific knowledge gap in the financial literature.

Consider D stocks that possess a certain relationship with each other (for instance, similar geographical locations, industry, competitors, etc.). For the sake of simplicity, let $D = 3$ and the stocks have a considerable linear relationship among each other to be able to easily visualize this stock data. Fig. 8 shows the 3D scatter plot of the normalized stock returns where each data point is represented by the vector $\boldsymbol{S}_t = \{S_{t,a},\ S_{t,b}, S_{t,c}\}$ at time t.

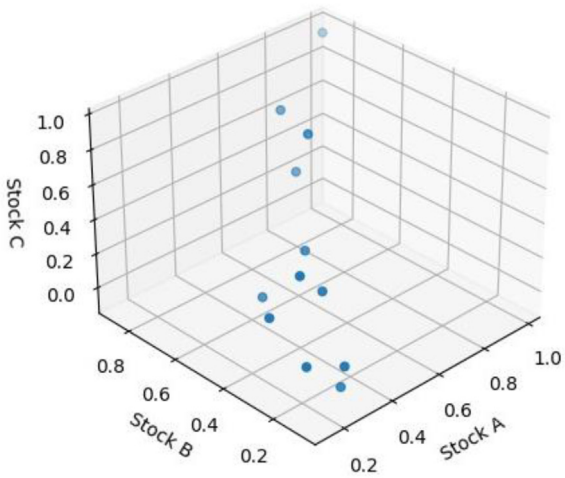

Fig. 8. 3D scatter plot for normal periods.

Also consider the distance metric $d(\boldsymbol{S}_t, \boldsymbol{S}_{t+\tau})$, where $\tau \in \mathbb{Z}/\{0\}$, and $\overline{d} := \boldsymbol{M}[d(\boldsymbol{S}_t, \boldsymbol{S}_{t+\tau})] \forall t \,\&\, \tau \in \mathrm{T}$ and $t \neq \tau$, where T is the set containing all considered time steps (for Fig. 7, $|\mathrm{T}| = 12$.) and $\boldsymbol{M}$ is the operator. Based on the theoretical and empirical findings related to Financial Tail Dependence Theory (Beine et al., 2010; Chesnay and Jondeau, 2001; White et al., 2015), which states that normally the absolute correlations between stocks considerably increase during financially turbulent periods, we know that $E[\overline{d}_N] > E[\overline{d}_T]$, where E is the expected value, $\overline{d}_N$ is $\overline{d}$ during a normal period and $\overline{d}_T$ is $\overline{d}$ during a turbulent period. In other words, since the absolute correlations between the stocks in $\boldsymbol{S}_t$ increase during turbulent periods, depicted by Fig. 9, the average distances between the data points will normally be smaller than the average distances during normal periods.

Given the findings of Jebran et al. (2017) that indicate that the absolute correlations between stocks also increase in periods preceding turbulent periods, it can be conjectured that $E[\overline{d}_P] \geq E[\overline{d}_T]$ and $E[\overline{d}_N] > E[\overline{d}_P]$, where $\overline{d}_P$ is $\overline{d}$ during periods preceding turbulent periods. Consequently, the following inequality holds:

$$|E[\overline{d}_N] - E[\overline{d}_P]| \geq |E[\overline{d}_T] - E[\overline{d}_P]| \tag{6}$$

$$|E[\overline{d}_N] - E[\overline{d}_P]| > |E[\overline{d}_{N1}] - E[\overline{d}_{N2}]| \tag{7}$$

where, $\overline{d}_{N1}$ is $\overline{d}$ during normal period 1 and $\overline{d}_{N1}$ is $\overline{d}$ during the subsequent normal period.

The presumable increase in absolute correlations found by Jebran et al. (2017) is depicted in Fig. 10.

Thus, the difference between the average distance of two periods can be used as an indicator to foresee a financially turbulent period by defining a threshold value to be used during normal periods. Nonetheless, the issue with this approach is the fact that $\overline{d}$ suffers from the curse of dimensionality and fails to detect nonlinear and complex relationships in the data (Pereira and De Mello, 2015; Shnier et al., 2019). The curse of dimensionality that $\overline{d}$ suffers is explained by the fact that as $D \to \infty, \frac{d(X_t, X_{t+\tau})}{d(X_t, X_{t+\psi})} \to 1$ (Pereira and De Mello, 2015). On the other hand, the implementation of PH information through WD or L^n norms of Persistent Landscape does not suffer from these issues (Pereira and De Mello, 2015; Shnier et al., 2019). Hence, this is the reason for the success of the implementation of PH information in recent studies and its choice in this study.

Incidentally, it is worth noting that although the aforementioned example makes use of only three dimensions/stocks and normalized stock returns to facilitate the comprehension of the proposed theory, the theory also works with a higher number of dimensions/stocks and non-normalized stock returns.

When observing the two crises (2008 and 2020 crises) present in the main sample data of this study, which is composed by the stock indexes Standard & Poor's 500 (S&P 500), Russell 2000 (RUT), and Dow Jones Industrial Average (DJIA), the predicted behavior in Topological Tail Dependence Theory can be seen. Figs. 11–13 respectively show 3D scatter plots of the returns of the main sample considering 22 business days from July 14, 2008 until August 12, 2008 (normal period), August 13, 2008 until September 12, 2008 (preceding period), and from September 15, 2008 until October 14, 2008 (turbulent period).

The correlations between the stocks in the main sample increase in the preceding period in comparison to the normal period, and they increase, yet to a smaller extent, in the turbulent period in comparison to the preceding period. Similarly, the same pattern can be seen in a more pronounced way for the 2020 crisis. Again, Figs. 14–16 respectively show 3D scatter plots of the returns of the main sample considering 22 business days from December 16, 2019 until January 16, 2020 (normal period), from January 17, 2020 until February 19, 2020 (Preceding Period) (preceding period), and February 20, 2020 until March 23, 2020 (turbulent period).

Finally, in the lecture "Seeking New Laws" on November 19, 1964, Richard Feynman said "if it (the consequences from your theory) disagrees with experiments, it is wrong". Thus, to test the proposed theory in this paper, its consequences must first be estimated. One

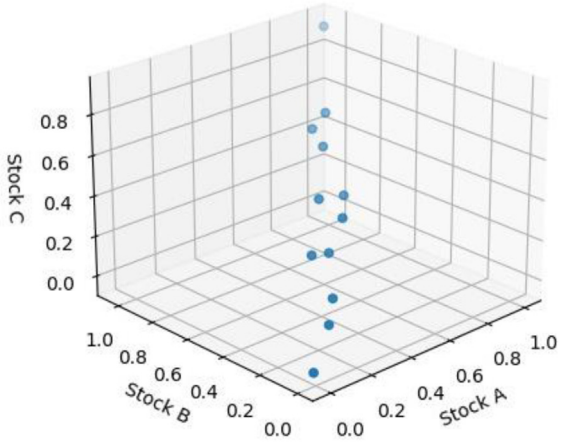

Fig. 9. 3D scatter plot for turbulent period.

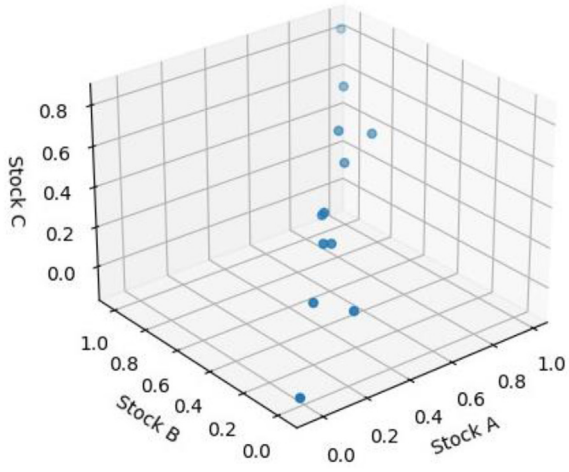

Fig. 10. 3D scatter plot for preceding period.

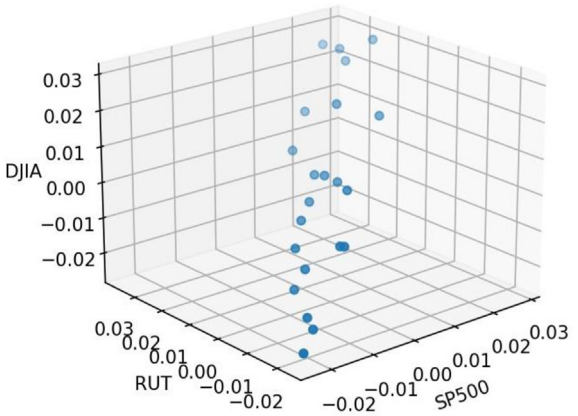

Fig. 11. 3D scatter plot from July 14, 2008 until August 12, 2008 (normal period).

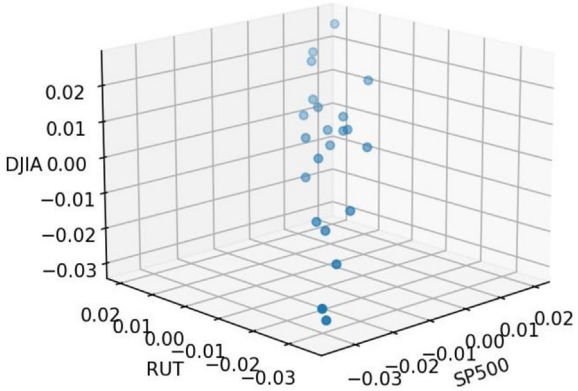

Fig. 12. 3D scatter plot from August 13, 2008 until September 12, 2008 (preceding period).

very natural consequence of Topological Tail Dependence Theory is the expectation that if PH information is used to forecast realized volatility, it should in any case improve the model's forecast performance during turbulent periods. This is the case since this additional information will allow the model to better foresee high values of stock realized volatility, which occur during turbulent periods. Hence, the remainder of this study is dedicated to empirically testing this expectation to put the proposed theory to the test.

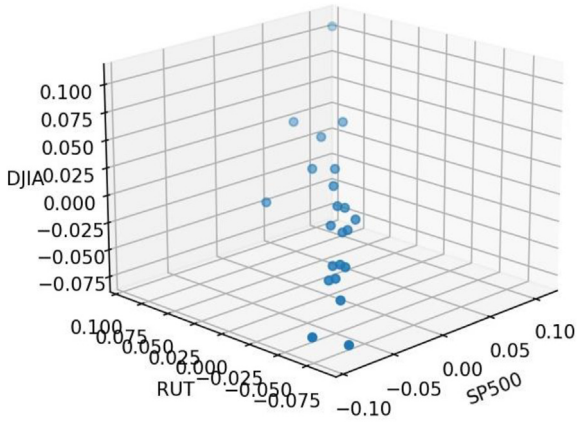

Fig. 13. 3D scatter plot from September 15, 2008 until October 14, 2008 (turbulent period).

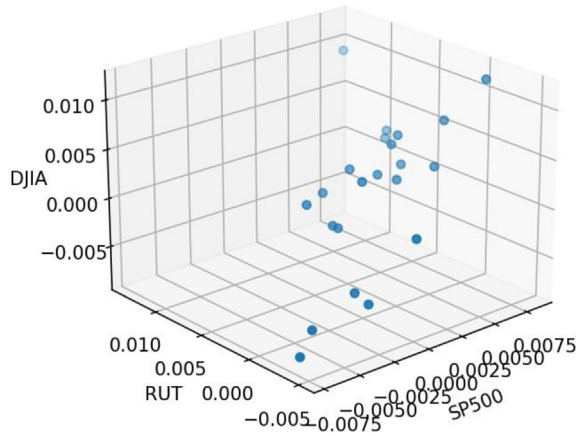

Fig. 14. 3D scatter plot from December 16, 2019 until January 16, 2020 (normal period).

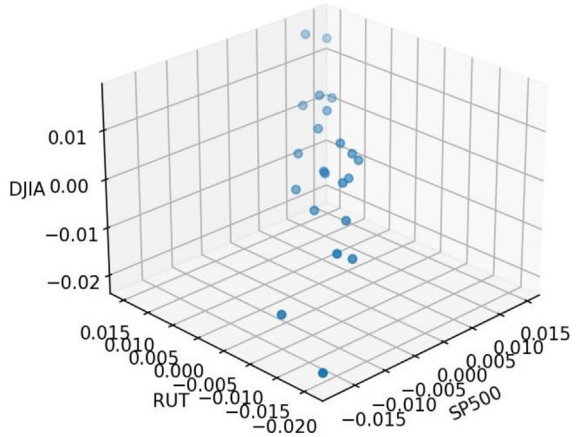

Fig. 15. 3D scatter plot from January 17, 2020 until February 19, 2020 (preceding period).

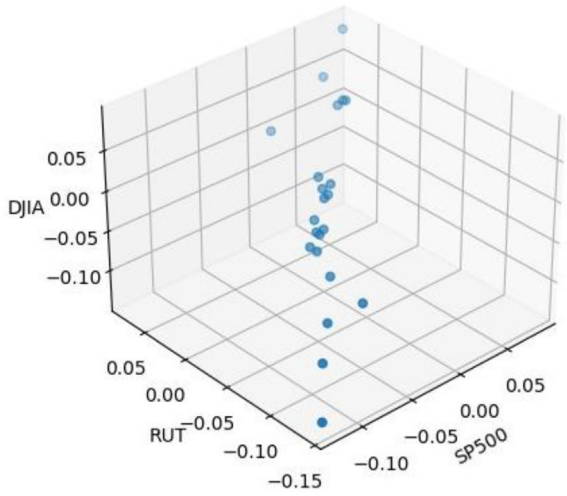

Fig. 16. 3D scatter plot from February 20, 2020 until March 23, 2020 (turbulent period).

4. Methodology

4.1. Sample

For the main sample, three famous stock indexes are used, namely Standard & Poor's 500 (S&P 500), Dow Jones Industrial Average (DJIA) and Russell 2000 (RUT). The sample period for these stock indexes ranges from the 1st of January 2000 until 30 March 2022. However, after employing the PH-RV algorithm, this period ranges from 1st of February 2000 until 2 March 2022, encompassing a total of 5557 daily observations. The choice of a large sample period is motivated by five key points present in the forecasting daily realized volatility literature, namely:

1. **Market Dynamics and Structural Changes:** The financial markets are known to be subject to structural changes over time, including shifts in volatility regimes, changes in market microstructure, and regulatory reforms. A long sample period allows for the incorporation of various market conditions, thereby capturing a broader range of market dynamics (Engle, 2002);
2. **Statistical Robustness:** A longer sample period provides a larger number of observations, which can enhance the statistical robustness of your analysis. This can lead to more precise parameter estimates and more stable model performance assessments (Elliott and Timmermann, 2008);
3. **Out-of-Sample Testing:** By extending the sample period, you can also incorporate more out-of-sample data for forecast evaluation. This helps to reduce the risk of overfitting and provides a more realistic assessment of how well your volatility forecasting model generalizes to unseen data (Hansen, 2005);
4. **Longer Economic and Financial Cycles:** Financial markets are influenced by longer economic and financial cycles, and a longer sample period allows for capturing these cycles. This is important for assessing the performance of volatility forecasts, as they need to adapt to changing economic conditions (Pagan and Schwert, 1990) and;
5. **Improved Forecast Stability:** Volatility forecasting models can be sensitive to the choice of sample period. Using a longer sample period can lead to more stable forecasts and reduce the impact of short-term noise in the data (Andersen et al., 2003).

As a result, a great variety of studies in this literature choose to have large sample sizes when possible (Corsi et al., 2008; Deo et al., 2006; Wang et al., 2016). Additionally, the exact start and end dates were chosen based on data availability in order to maximize the number of considered data points. The same sample is used for two out of the three robustness tests discussed in Section 3.6. The third robustness test, on the other hand, employs the following three European stock indexes: Financial Times Stock Exchange 100 (FTSE), DAX Performance (GDAXI), and STXE 600 (STOXX). For these European indexes, the sample period after the application of the PH-RV algorithm ranges from 25 May 2004 until 2 March 2022, a total of 4390 daily observations. The sample period for the European indexes is a bit shorter due to data availability.

Further, both samples are divided into 80% training sample and 20% test sample. Moreover, while the training sample is used to estimate the optimal parameters of the studied models, 25% of the training sample is used as validation sample to determine the hyperparameters of the neural network models (see Section 3.4 for more details). Additionally, in order to test the Topological Tail Dependence Theory, a subsample of the test sample representing the 2020 crisis will also be used. This subsample ranges from 20 February 2020 until 21 May 2020.

Lastly, all stock data in this research is retrieved from TAQ and eurofidai dataset.

4.2. Realized volatility proxy

The RV proxy chosen for this paper is Yang & Zhang's RV (2000) due to its robustness to drift and consistency in dealing with opening price jumps (Yang and Zhang, 2000), yielding statistically more reliable and robust realized volatility measures than the normal technique of summing the intraday squared returns (Alizadeh et al., 2002; Tsatekidis et al., 2017; Yarovaya et al., 2016). Although Yang and Zhang (2000) originally applied their model only with daily prices to estimate weekly, monthly, or yearly volatility, their volatility proxy can be easily adapted to daily realized volatility by making use of intraday prices instead of daily prices. In this research, 5-min interval High, Low, Open and Close prices are used. Mathematically, Yang & Zhang's RV is defined as:

$$RV_t = \sqrt{\sigma_t^2} \tag{8}$$

$$\sigma_t^2 = \sigma_{O_t}^2 + k\,\sigma_{C_t}^2 + (1-k)\sigma_{RS_t}^2 \tag{9}$$

with,

$$\sigma_{O_t}^2 = \frac{1}{M-1}\sum_{i=1}^{M}(o_i - \overline{o}) \tag{10}$$

$$\sigma_{C_t}^2 = \frac{1}{M-1}\sum_{i=1}^{M}(c_i - \overline{c}) \tag{11}$$

where, o_i = 5-mintue interval opening price at time i of the t-th day, $\overline{o}$ = 5-min interval opening price mean of the t-th day, c_i = 5-min interval close price at time i of the t-th day, $\overline{c}$ = 5-min interval close price mean of the t-th day, k = parameter, M = total number of 5-min intervals on the t-th day, and $\sigma_{RS_t}^2$ = Rogers and Satchell (1991) variance estimation of the t-th day also using 5-min interval prices.

Yang and Zhang's (2000) empirical research indicates that the best k value is given as:

$$k = \frac{0.34}{1.34 + \frac{M+1}{M-1}} \tag{12}$$

where, M = total number of 5-min intervals on the t-th day. Therefore, this k value is used in this research under the assumption that it also holds for intraday data.

4.3. Considered models

This paper examines two neural network models, two nonlinear models and two linear models. Given the recent yet impressive superiority of Neural Basis Expansion Analysis with incorporate Exogenous factors (NBEATSx) over other state-of-the-art neural network models in both generic forecasting tasks (Olivares et al., 2022; Oreshkin et al., 2021; Sbrana et al., 2020; Wang et al., 2022) and forecasting realized volatility (Souto and Moradi, 2023a) and covariance matrix (Souto and Moradi, 2023b), the choice of NBEATSx comes naturally. Therefore, the two neural network models are NBEATSx without any exogenous variable (i.e., NBEATS) and NBEATSx with 2-d Wasserstein distances (2DWD) estimated by the PH-RV algorithm. The first model is henceforth mentioned as NBEATS while the latter model as NBEATSx-PH. Additionally, for a thorough explanation of NBEATSx architecture, see Olivares et al. (2022) and Souto and Moradi (2023).

Concerning the linear models, the standard benchmarks Heterogeneous Autoregressive (HAR) model (Corsi, 2009) and Heterogeneous Autoregressive with eXogenous variables (HARX) model (Hillebrand and Medeiros, 2010) are employed. The HAR and HARX models are respectively defined by Equations (13) and (14):

$$\widehat{RV_t} = \beta_0 + \beta_d RV_{t-1} + \beta_w \overline{RV}_{t-1:t-5} + \beta_m \overline{RV}_{t-1:t-22} \tag{13}$$

$$\widehat{RV_t} = \beta_0 + \beta_d RV_{t-1} + \beta_w \overline{RV}_{t-1:t-5} + \beta_m \overline{RV}_{t-1:t-22} + \boldsymbol{\beta_x}\,\boldsymbol{X_{t-1}} \tag{14}$$

where, $\widehat{RV_t}$ is the forecasted RV_t, $\overline{RV}_{t-1:t-5}$ and $\overline{RV}_{t-1:t-22}$ are respectively the weekly and monthly moving average of RV and $\boldsymbol{X_{t-1}}$ is the $1 \times N$ vector of lagged exogenous variables, with N being the number of considered exogenous variables. For this study, the HARX model encompasses 2DWD as an exogenous variable.

A natural question that may arise from the choice of the HAR model as the linear model used in this research is: why choose the HAR model instead of one of the famous GARCH models? The choice of the HAR model is motivated by the following four facts:

1. **Long Memory in Volatility:** Empirical evidence suggests that financial time series data often exhibit long-memory volatility patterns, where past volatility observations have a persistent impact on future volatility (Corsi, 2009). The HAR model incorporates multiple past lags of volatility, capturing this long-memory property more effectively (Corsi, 2009).

2. **Higher Forecasting Accuracy:** Studies comparing HAR and GARCH models have consistently shown that HAR models outperform GARCH models in terms of forecasting realized volatility. They tend to provide more accurate and reliable forecasts, particularly in high-frequency financial data (Corsi, 2009; Hillebrand and Medeiros, 2010).
3. **Incorporating Information from Different Frequencies:** HAR models consider volatility information at multiple frequencies, including daily, weekly, and monthly. This helps in capturing the effects of various market dynamics that affect volatility differently over time, making them more adaptable to changing market conditions (Corsi et al., 2010).
4. **Robustness:** HAR models are relatively robust to various data sampling frequencies and are capable of handling irregularly spaced data, which is common in financial markets. This robustness enhances their applicability in different financial forecasting scenarios (Patton and Sheppard, 2015).

Lastly, the two nonlinear models chosen for this study are the famous Multiple-Regime Smooth Transition Heterogeneous Autoregressive (HARST) model (McAleer and Medeiros, 2008) with the weekly moving average of RV and 2DWD as transition variables. The HARST model as proposed by McAleer and Medeiros (2008) is defined as:

$$\widehat{RV_t} = \beta_0 + \beta_d RV_{t-1} + \beta_w \overline{RV}_{t-1:t-5} + \beta_m \overline{RV}_{t-1:t-22} + \sum_{m=1}^{M} \sum_{b=1}^{3} \widetilde{RV_b} \beta_{m,b} f(z_t; \gamma_m, c_m) \tag{15}$$

where, $\widetilde{RV_1} = RV_{t-1}$, $\widetilde{RV_2} = \overline{RV}_{t-1:t-5}$, $\widetilde{RV_1} = \overline{RV}_{t-1:t-22}$, and $f(z_{t-1}; \gamma_m, c_m)$ is given as:

$$f(z_{t-1}; \gamma_m, c_m) = \frac{1}{1 + e^{-\gamma_m (z_{t-1} - c_m)}} \tag{16}$$

where, z_{t-1} is the transition variable at time $t-1$, γ_m and c_m are parameters for the m-th regime. The main advantage of the HARST model is that it can capture both long-range dependence and regime switches (and hence asymmetric effects) in a very simple way. For instance, if the transition variable is stock return at time $t-1$, then two regimes ($M = 2$) can be constructed with one for when the stock return at time $t-1$ is positive and the other when it is negative. As a result, the differences in the dynamics of the conditional variance are modelled according to the sign of the shocks in previous returns, which represent previous "news".

In the case of the two nonlinear models used in this research, in accordance with the findings in McAleer and Medeiros (2008), the best number of regimes found is equal to two periods and the transition threshold for both transition variables is when the variable value surpasses the 75th historical percentile. By employing this threshold, the model switches from a normal-period regime to a turbulent-period regime for weekly moving average of RV as a transition variable and switches from a normal-period regime to a preceding-turbulent-period and turbulent-period regime for 2DWD as a transition variable under the assumption that the Topological Tail Dependence Theory is indeed correct. Additionally, under the assumption that the Topological Tail Dependence Theory is indeed correct, it is expected that the latter model will have a superior forecast performance (at least during turbulent periods).

The choice for the HARST model as the nonlinear model used in this research is motivated by its popularity in the realized volatility forecast literature (McAleer and Medeiros, 2010), its superior and robustness forecast performance with respect to other models (Corsi et al., 2012), and its simplicity (Buccheri and Corsi, 2019). Finally, the parameters of the linear models are estimated through Ordinary Least Squares (OLS), while the parameters of the nonlinear models and the neural network models are respectively estimated through the Quasi-Likelihood method proposed by McAleer and Medeiros (2008) and minimizing the chosen loss function.

4.4. Hyperparameters search space

Table 1 presents the hyperparameters search space utilized to determine the optimal hyperparameters for each studied stock index using the validation sample results.

A total of 65 trials are performed for each stock index to explore the hyperparameter search space, of which 20 trials are specially dedicated to exploring the random seed range.

Table 1
Hyperparameters search space.

Hyperparameters	NBEATSx
Number of Lags	[5, 10, 21, 63, 84, 189, 252]
Scaler Type	[robust, standard, minmax]
Number of Epochs	[25, 50, 100, 150, 250, 350, 450, 550, 750]
Loss Functions	[MSE, MAE, MQLoss (level = [90]), MQLoss (level = [80, 90]), MQLoss (level = [95]), MQLoss (level = [75])]
Learning Rate	[5.00E^{-4}, 1.00E^{-4}, 5.00E^{-5}, 1.00E^{-5}]
Number of Learning Decays	[5, 3, 2, 1]
Number of units of each hidden layer	[[[712, 712], [712, 712]], [[512, 512], [512, 512]], [[250, 250], [250, 250]], [[100, 100], [100, 100]]]
Number of blocks per stack	[[1, 1], [2, 2], [3, 3], [5, 5]]
Stack Types	[[identity, identity, exogenous], [identity, trend, exogenous], [identity, seasonality, exogenous], [trend, seasonality, exogenous]]
Number of Harmonics	[0, 1]
Number of Polynomials	[0, 1]
Random Seed Range	[0, …, 129228148]

4.5. Forecast assessment

This paper employs three error metrics to assess the models' forecast accuracy in the out-sample data (test sample). These are namely Root-Mean-Square Error (RMSE), Mean Absolute Error (MAE), and Quasi-likelihood (QLIKE). RMSE and MAE are classical error measures used to assess forecast performance of models for forecasting realized volatility (Corsi et al., 2008; Deo et al., 2006; Dutta and Das, 2022; Gallo and Otranto, 2015; Wang et al., 2016; and many others). The choice of RMSE is also motivated by its easily interpretable results when compared to Mean-Square Error (MSE) while being more sensitive to errors further from the mean than MAE (Chiriac and Voev, 2010; Corsi et al., 2008; Dutta and Das, 2022). QLIKE, on the other hand, is utilized to account for errors in a relative manner (i.e., QLIKE is a scale-free error measure) (Cipollini et al., 2020; Patton and Sheppard, 2009; Zhang et al., 2023). This property of QLIKE allows it to be easily used to compare model's performance in datasets with different scales. Concerning RV, QLIKE can be used to compare a model's accuracy performance among stock indexes of high and low volatility. The selected error measures are explained by Equations (17)–(19):

$$RMSE = \sqrt{\frac{\Sigma_{t=1}^{T} \left(RV_t - \widehat{RV_t}\right)^2}{T}} \tag{17}$$

$$MAE = \frac{\Sigma_{t=1}^{T} \left|RV_t - \widehat{RV_t}\right|}{T} \tag{18}$$

$$QLIKE = \frac{\Sigma_{t=1}^{T} \left[\frac{RV_t}{\widehat{RV_t}} - \log\left(\frac{RV_t}{\widehat{RV_t}}\right) - 1\right]}{T} \tag{19}$$

where, T is the number of forecasted time-steps and $\widehat{RV_t}$ is the forecasted RV_t. Additionally, since RV values are always positive, $\log\left(\frac{RV_t}{\widehat{RV_t}}\right)$ is always well defined.

Further, Diebold-Mariano (DM) tests are performed to determine whether the forecasts from a pair of models present statistically significantly different accuracy. DM tests were originally proposed by Diebold and Mariano (1995), and later modified by Harvey et al. (1997) to take serial dependence into account. DM tests are also widely used in the realized volatility forecast literature (Dutta and Das, 2022; Gallo and Otranto, 2015; Zhang et al., 2023). Consequently, this paper employs the modified DM tests as proposed by Harvey et al. (1997) with RMSE, MAE and QLIKE as the error metrics. The following two hypotheses are used for DM tests:

- Null Hypothesis (H_0): there is no statistically significant difference between the forecasts of the considered pair of models.
- Alternative Hypothesis (H_1): there is a statistically significant difference between the forecasts of the considered pair of models.

In addition, Model Confidence Set (MCS) tests, proposed by Hansen et al. (2011), are also employed in this research. MCS tests identify a subset of models with significantly superior performance from model candidates, at a given level of confidence (a p-value threshold). For more details about this procedure, see Hansen et al. (2011). Similar to DM tests, MCS tests are also common in the realized volatility forecast literature (Chiriac and Voev, 2010; Gallo and Otranto, 2015; Zhang et al., 2023). Incidentally, MCS tests are also performed with RMSE, MAE and QLIKE as the error metrics.

In this research, the used p-value thresholds are 0.01, 0.05, and 0.10, and they are respectively represented in the tables of Section 5 and 6 by ***, **, and *. DM tests are only performed for the pair of models in the same "family" (i.e., for the considered pair of neural network models, for the pair of nonlinear models, and for the pair of linear models) in order to measure the effect of the implementation of PH information for forecasting RV in a statistical manner.

Finally, graphical comparisons are performed for the forecasts of each stock index. However, to ensure sparsity, only the graphical comparison of S&P 500 for the main sample is displayed in this paper. To find the graphical comparisons for all stock indexes and robustness tests, please see https://github.com/hugogobato/Topological-Tail-Dependence-Evidence-from-Forecasting-Realized-Volatility.git.

4.6. Robustness tests

Three aspects regarding the robustness of the proposed model are considered: (i) an alternative training set size for the main sample, (ii) alternative data splitting, and (iii) stock indexes from other geographic regions than the main sample.

For the first robustness test (i), the training sample size is reduced to half of its original size, while for the second robustness test (ii), the data splitting is 60%/40% in contrast to 80%/20% in the main sample. Finally, the third robustness test (iii) employs European stock indexes, namely FTSE, GDAXI, and STOXX.

5. Main sample results

Table 2 shows the chosen optimal hyperparameters of the neural network models for each index after the validation stage

Table 3 presents the estimated coefficients of the HARX models. Interestingly, the coefficients for 2DWD are all statistically significant at the level of 0.01, showing that 2DWD could potentially have a significant linear relationship with RV. However, in order to confirm this, a similar linear regression ought to be done with control variables and the forecasts of the HARX models should be superior to the forecasts of the HAR models (at least during turbulent periods based on Topological Tail Dependence Theory).

Table 4 also presents the estimated coefficients of the HARX models, but now with the incorporation of the control variables VIX close prices and Fama-French's three factors. Incidentally, the motivation behind the choice of such control variables can be found in Appendix A. Similar to the results of Table 3, the coefficients for 2DWD are all statistically significant, but now at the level of 0.01 for S&P 500 and DJIA, and at the level of 0.05 for RUT. Although these results increase the likelihood that 2DWD has a significant linear relationship with RV, it is still necessary to check whether the forecasts of the HARX models are indeed superior to the forecasts of the HAR models (at least during turbulent periods based on Topological Tail Dependence Theory).

Table 5, on the other hand, presents the error measures and *p*-values of the statistical tests for each index concerning the full sample. It is worth recalling that the *p*-values shown for the DM test only concern the pair of models in the same model family (i.e., NBEATS vs NBEATSx-PH, HAR vs HARX, and HARST-week vs HARST-PH). At first glance, NBEATSx-PH is clearly the best model among all other considered models, which is not surprising given the flexibility and power of neural network models. When comparing NBEATS with NBEATSx-PH, it is clear that the second model is statistically significantly superior, with the exception of RUT for the error measure QLIKE. On the other hand, the results for the comparison between the HAR and HARX models are not as clear. Although the HARX model is systematically better than HAR regarding RMSE and QLIKE, this superiority is not statistically significant, with the exception of DJIA. Similarly, even though HARST-PH systematically yields more accurate forecasts than HARST-week concerning all error measures, this superiority is not statistically significant, with the exception of RUT for the error measure QLIKE.

However, when considering the forecast results for the 2020 crisis, found in Table 6, the models that incorporate PH information systematically yield statistically more precise forecasts for at least one error measure. This is exactly what the Topological Tail Dependence Theory predicted. Seemingly, the incorporation of PH information allows the models to better foresee turbulent periods. Needless to say, this advantage is more expressed in more flexible models than rigid models. Additionally, NBEATSx-PH is anew clearly the best model among all other models.

Fig. 17 displays the graphical comparison for the forecasts for S&P 500. Although the linear and nonlinear models are seemingly better than the neural network models, given the results in Table 6, this is probably a graphical illusion due to a great number of data points shown in the graph and its resolution. When looking at Fig. 18, it is clear that the neural network models perform better during turbulent periods.

Despite the results for the main sample failing to disproof the Topological Tail Dependence Theory, robustness tests are still required to conclude whether the Topological Tail Dependence Theory is indeed a promising theory that ought to be explored further.

6. Robustness tests

Once again, the robustness tests performed and discussed in this section are (i) an alternative training set size for the main sample, (ii) the data splitting is 60%/40% in contrast to 80%/20% in the main sample, and (iii) stock indexes from other geographic regions than the main sample.

Table 2
Optimal Hyper parameters.

Models	Hyperparameters		
	S&P 500	DJIA	RUT
NBEATS	Number of lags: 63 Scaler Type: minmax Stack Types: [identity, identity] Number of units of each hidden layer: [[712, 712],[712, 712]] Learning rate: $1.00E^{-4}$ Loss Function: MQLoss (level = [90]) Number of blocks: [2, 2] Number of epochs: 250 Random seed: 67671771	Number of lags: 63 Scaler Type: robust Stack Types: [identity, identity] Number of units of each hidden layer: [[512, 512],[512, 512]] Learning rate: $5.00E^{-4}$ Loss Function: MQLoss (level = [75]) Number of blocks: [5, 5] Number of epochs: 100 Random seed: 83642934	Number of lags: 63 Scaler Type: minmax Stack Types: [identity, identity] Number of units of each hidden layer: [[712, 712],[712, 712]] Learning rate: $1.00E^{-5}$ Loss Function: MAE Number of blocks: [2, 2] Number of epochs: 750 Random seed: 22782732
NBEATSx-PH	Number of lags: 21 Scaler Type: minmax Stack Types: [identity, identity, exogenous] Number of units of each hidden layer: [[250, 250],[250, 250]] Learning rate: $1.00E^{-4}$ Loss Function: MSE Number of blocks: [1, 1] Number of epochs: 550 Random seed: 69871046	Number of lags: 5 Scaler Type: standard Stack Types: [identity, identity, exogenous] Number of units of each hidden layer: [[712, 712],[712, 712]] Learning rate: $1.00E^{-5}$ Loss Function: MAE Number of blocks: [5, 5] Number of epochs: 150 Random seed: 62404816	Number of lags: 5 Scaler Type: standard Stack Types: [identity, identity, exogenous] Number of units of each hidden layer: [[512, 512],[512, 512]] Learning rate: $1.00E^{-5}$ Loss Function: MQLoss (level = [95]) Number of blocks: [1, 1] Number of epochs: 450 Random seed: 97819447

Table 3
HARX models coefficients.

Coefficients	Measures	S&P 500	DJIA	RUT
β_0	Coefficient	0.0004	0.0005	0.0007
	T-statistics *p*-value	3.65 2.77E^{-4}***	3.91 9.70E^{-5}***	4.69 3.06E^{-6}***
2DWD	Coefficient	0.018	0.016	0.015
	T-statistics *p*-value	6.81 1.39E^{-11}***	6.37 2.53E^{-10}***	5.34 1.12E^{-7}***
$\overline{RV}_{t-1:t-22}$	Coefficient	0.227	0.223	0.249
	T-statistics *p*-value	7.78 1.34E^{-14}***	7.52 9.67E^{-14}***	8.30 3.09E^{-16}***
$\overline{RV}_{t-1:t-5}$	Coefficient	0.501	0.522	0.456
	T-statistics *p*-value	15.57 1.34E^{-50}***	16.15 9.28E^{-54}***	14.14 6.85E^{-42}***
RV_{t-1}	Coefficient	0.135	0.118	0.161
	T-statistics *p*-value	7.45 1.62E^{-13}***	6.49 1.19E^{-10}***	8.90 2.18E^{-18}***

Table 4
HARX models coefficients with control variables.

Coefficients	Measures	S&P 500	DJIA	RUT
β_0	Coefficient	−0.001	−0.0008	0.0002
	T-statistics *p*-value	−7.61 6.93E^{-14}***	−5.24 2.03E^{-7}***	1.35 0.17
2DWD	Coefficient	0.018	0.015	0.006
	T-statistics *p*-value	7.26 8.51E^{-13}***	6.03 2.42E^{-9}***	2.12 0.03**
$\overline{RV}_{t-1:t-22}$	Coefficient	−0.130	−0.078	0.122
	T-statistics *p*-value	−3.64 2.88E^{-4}***	−2.18 0.02**	3.82 1.48E^{-4}***
$\overline{RV}_{t-1:t-5}$	Coefficient	0.400	0.445	0.443
	T-statistics *p*-value	12.69 4.63E^{-34}***	14.04 2.62E^{-40}***	14.20 3.17E^{-40}***
RV_{t-1}	Coefficient	0.044	0.042	0.101
	T-statistics *p*-value	2.57 0.01**	2.39 0.01**	5.83 8.70E^{-9}***
VIX	Coefficient	0.0003	0.0002	0.0001
	T-statistics *p*-value	15.70 2.90E^{-49}***	13.85 2.43E^{-39}***	9.37 1.06E^{-19}***
Mkt-RF	Coefficient	−0.0006	−0.0006	−0.0008
	T-statistics *p*-value	−12.86 6.97E^{-35}***	−12.64 1.43E^{-33}***	−15.50 1.38E^{-46}***
SMB	Coefficient	−0.0004	−3.43E^{-4}	−0.0007
	T-statistics *p*-value	−4.43 1.08E^{-5}***	−3.84 1.30E^{-4}***	−6.11 1.69E^{-09}***
HML	Coefficient	6.31E^{-5}	7.92E^{-5}	0.0001
	T-statistics *p*-value	0.78 0.43	−0.98 0.32	1.41 0.15

6.1. Robustness test (i)

Table 7 shows the coefficients of the HARX models for the robustness test (i). Although the coefficients for the lags of RV vary to a great extent, the coefficients for the intercept and 2DWD do not. Additionally, the coefficients for 2DWD are still statistically significant at the level of 0.05 for all stock indexes, similar to the results of the main sample.

The coefficients of the HARX models with the incorporation of the control variables can be seen in Table 8. Similar to the results of Table 4, the coefficients for 2DWD are all statistically significant at the level of 0.01 for all stock indexes. These results further increase the likelihood that 2DWD has a significant linear relationship with RV.

Table 9, on the other hand, presents the forecasting results for the full test sample of the robustness test (i). Anew, NBEATSx-PH is statistically superior to other models for all stock indexes and error measures. Similar to the results in Table 5, the incorporation of PH information systematically improved the forecast performance of all model families. However, this time, the improvement is statistically significant for all models, apart from the HAR model for RUT. The forecast accuracy improvement being more present in nonlinear and flexible models can be explained by the Topological Tail Dependence Theory since based on this theory, the real value added of the incorporation of PH information only occurs during turbulent periods.

The forecasting results for the 2020 crisis can be found in Table 10. Similar to the main sample results, the incorporation of PH information systematically and statistically significantly increases the forecast accuracy of the considered model for at least one error measure per stock index. This provides further evidence that the implication/prediction of the Topological Tail Dependence Theory is empirically fulfilled. Again, NBEATSx-PH is the best model among all other models considering all error measures and stock indexes.

6.2. Robustness test (ii)

The estimated coefficients of the HARX models for the robustness test (ii) can be found in Table 11. Similar to the results of the main sample, the coefficients of 2DWD are statistically significant for all stock indexes at the level of 0.01. Likewise, the coefficients remain statistically significant for all stock indexes, but now at the level of 0.05, when adding the control variables as shown in Table 12. Thus, this further provides evidence for a possible significant linear relationship between RV and 2DWD.

The results for the full test sample for the robustness test (ii) can be found in Table 13. This time, NBEATSx-PH is not the best model considering QLIKE as an error measure. Regarding this error measure, the HARX model is superior to NBEATSx-PH. Additionally,

Table 5
Full sample results.

Stock Index	Measures	NBEATS	NBEATSx-PH	HAR	HARX	HARST-Week	HARST-PH
	RMSE	0.446%	**0.405%**	0.461%	0.455%	0.463%	0.448%
	MAE	0.275%	**0.258%**	0.292%	0.293%	0.292%	0.288%
S&P 500	QLIKE	8.80%	**8.13%**	9.99%	9.78%	9.96%	9.91%
	DM test (RMSE)	2.69E^{-3}***	2.69E^{-3}***	0.10	0.10	0.11	0.11
	DM test (MAE)	7.80E^{-5}***	7.80E^{-5}***	0.60	0.60	0.25	0.25
	DM test (QLIKE)	3.36E^{-4}***	3.36E^{-4}***	0.14	0.14	0.57	0.57
	MCS test (RMSE)	3.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (MAE)	1.00E^{-3}***	**1.00**	>1.00E^{-3}***	1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (QLIKE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	RMSE	0.509%	**0.429%**	0.528%	0.512%	0.530%	0.508%
	MAE	0.284%	**0.257%**	0.305%	0.305%	0.306%	0.302%
DJIA	QLIKE	8.73%	**7.67%**	9.89%	9.65%	9.94%	9.78%
	DM test (RMSE)	0.03**	0.03**	0.04**	0.04**	0.12	0.12
	DM test (MAE)	1.83E^{-5}***	1.83E^{-5}***	0.96	0.96	0.17	0.17
	DM test (QLIKE)	8.21E^{-5}***	8.21E^{-5}***	0.03**	0.03**	0.13	0.13
	MCS test (RMSE)	5.00E^{-3}***	**1.00**	4.00E^{-3}***	4.00E^{-3}***	4.00E^{-3}***	5.00E^{-3}***
	MCS test (MAE)	3.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (QLIKE)	2.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	RMSE	0.462%	**0.423%**	0.493%	0.487%	0.491%	0.479%
	MAE	0.311%	**0.293%**	0.342%	0.342%	0.343%	0.338%
RUT	QLIKE	7.94%	**7.69%**	9.08%	8.95%	9.15%	8.93%
	DM test (RMSE)	1.27E^{-3}***	1.27E^{-3}***	0.05*	0.05*	0.05*	0.05*
	DM test (MAE)	1.05E^{-3}***	1.05E^{-3}***	0.62	0.62	0.05*	0.05*
	DM test (QLIKE)	0.25	0.25	0.09*	0.09*	2.59E^{-3}***	2.59E^{-3}***
	MCS test (RMSE)	0.01**	**1.00**	1.00E^{-3}***	1.00E^{-3}***	1.00E^{-3}***	1.00E^{-3}***
	MCS test (MAE)	0.01**	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (QLIKE)	0.18	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***

The *p*-values shown for the DM test are only concerning the pair of models in the same model family (i.e., NBEATS vs NBEATSx-PH, HAR vs HARX, and HARST-week vs HARST-PH). Additionally, the best values out of all models are in bold to facilitate the read of the table.

Table 6
Results for the 2020 crisis.

Stock Index	Measures	NBEATS	NBEATSx-PH	HAR	HARX	HARST-Week	HARST-PH
	RMSE	1.137%	**0.934%**	1.183%	1.137%	1.206%	1.081%
	MAE	0.762%	**0.643%**	0.821%	0.821%	0.876%	0.876%
S&P 500	QLIKE	7.38%	**5.17%**	10.71%	7.61%	11.03%	8.38%
	DM test (RMSE)	0.02**	0.02**	0.04**	0.04**	0.04**	0.04**
	DM test (MAE)	0.03**	0.03**	0.08*	0.08*	0.05*	0.05*
	DM test (QLIKE)	9.60E^{-3}***	9.60E^{-3}***	0.08*	0.08*	1.60E^{-3}***	1.60E^{-3}***
	MCS test (RMSE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (MAE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (QLIKE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	RMSE	1.464%	**1.065%**	1.517%	1.478%	1.528%	1.393%
	MAE	0.861%	**0.658%**	1.010%	0.968%	1.019%	0.921%
DJIA	QLIKE	9.89%	**5.24%**	11.91%	9.15%	12.07%	8.86%
	DM test (RMSE)	0.08*	0.08*	0.03**	0.03**	0.10	0.10
	DM test (MAE)	8.18E^{-3}***	8.18E^{-3}***	0.06*	0.06*	0.06*	0.06*
	DM test (QLIKE)	3.46E^{-3}***	3.46E^{-3}***	0.05*	0.05*	2.46E^{-3}***	2.46E^{-3}***
	MCS test (RMSE)	1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (MAE)	1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (QLIKE)	1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	RMSE	0.969%	**0.717%**	1.038%	0.993%	1.009%	0.927%
	MAE	0.691%	**0.532%**	0.808%	0.770%	0.785%	0.732%
RUT	QLIKE	6.08%	**4.17%**	8.59%	6.41%	8.27%	6.12%
	DM test (RMSE)	8.26E^{-3}***	8.26E^{-3}***	0.04**	0.04**	0.08*	0.08*
	DM test (MAE)	0.01**	0.01**	0.12	0.12	0.14	0.14
	DM test (QLIKE)	0.01**	0.01**	0.01**	0.01**	5.42E^{-3}***	5.42E^{-3}***
	MCS test (RMSE)	2.00E^{-3}***	**1.00**	2.00E^{-3}***	2.00E^{-3}***	>1.00E^{-3}***	2.00E^{-3}***
	MCS test (MAE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (QLIKE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***

The *p*-values shown for the DM test are only concerning the pair of models in the same model family (i.e., NBEATS vs NBEATSx-PH, HAR vs HARX, and HARST-week vs HARST-PH). Additionally, the best values out of all models are in bold to facilitate the read of the table.

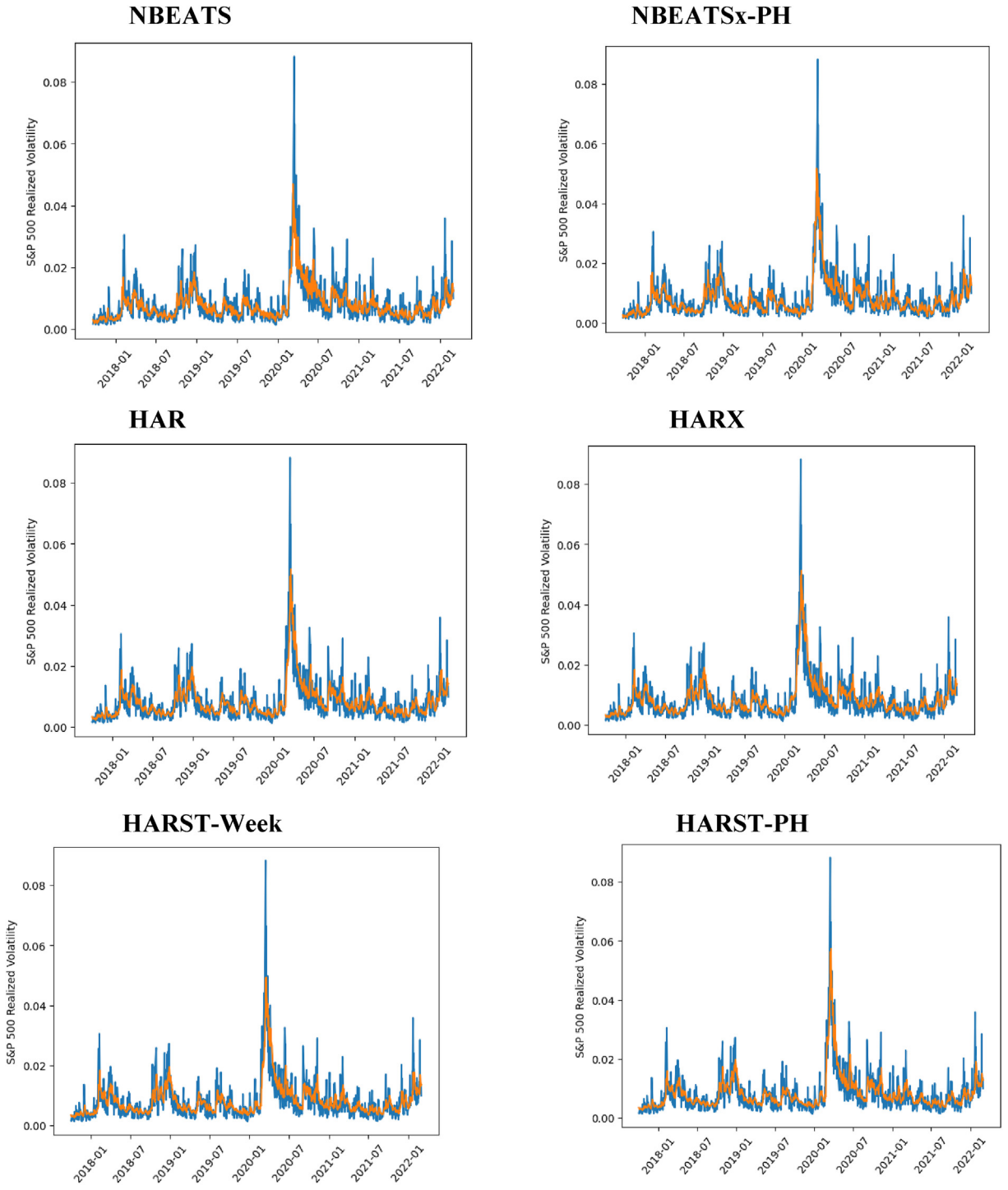

Fig. 17. Graphical comparison for S&P 500.

different to the results in Table 9, the forecasts of the HAR and HARX models are not statistically significantly different for any stock index when considering the *p*-value threshold of 0.05. However, apart from MAE, the incorporation of PH information systematically improves the forecast of the HAR model. Furthermore, for all stock indexes, the neural network and nonlinear models present a statistically significant forecast performance improvement through the incorporation of PH information for at least one error measure. Anew, the forecast accuracy improvement being more present in nonlinear and flexible models can be explained by the Topological Tail

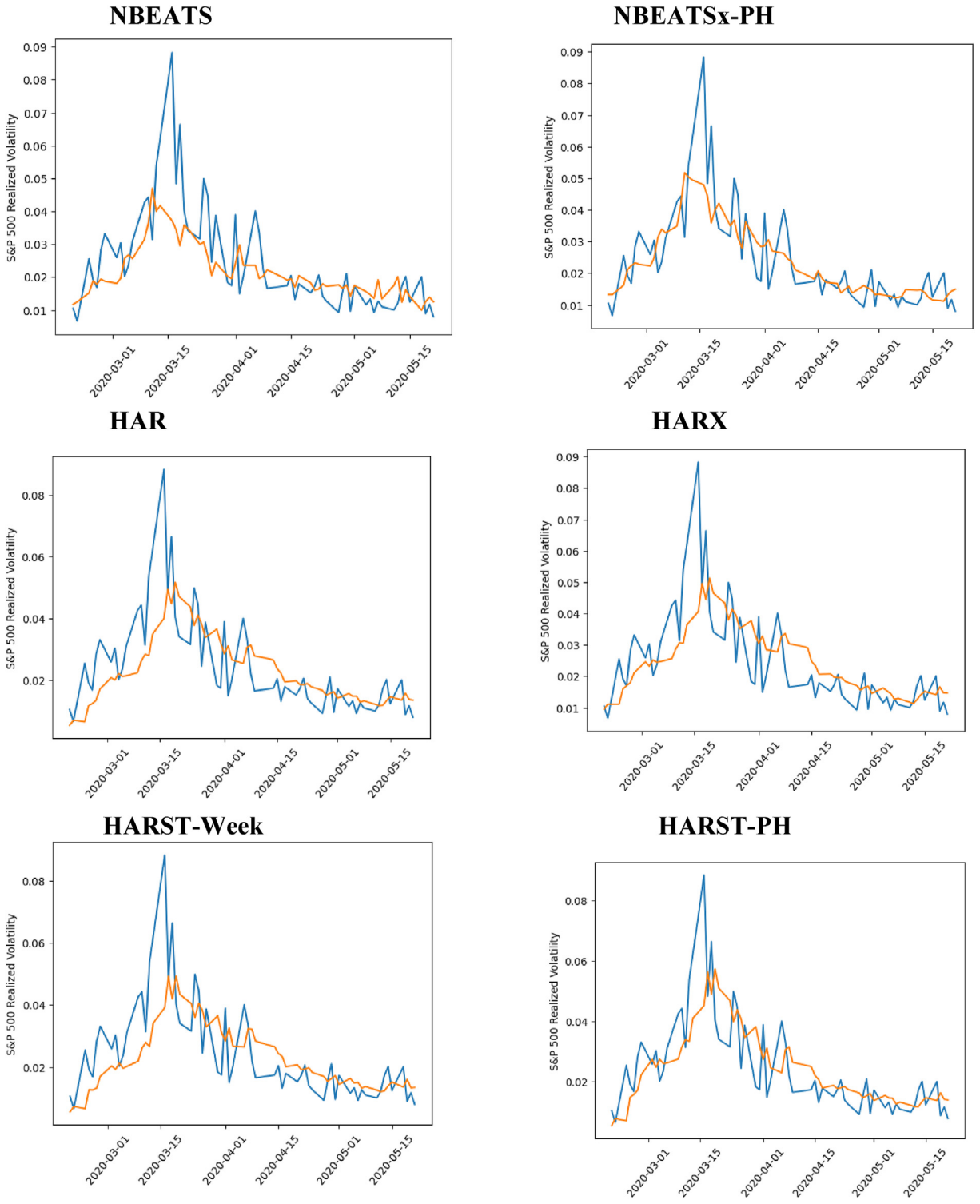

Fig. 18. Graphical comparison for S&P 500 during the 2020 crisis.

Dependence Theory since based on this theory, the real value added of the incorporation of PH information only occurs during turbulent periods.

Table 14, on the other hand, shows the forecasting results for the 2020 crisis. Again, as predicted by the Topological Tail Dependence Theory, the incorporation of PH information statistically significantly improves the forecast accuracy of all models for all stock indexes

Table 7
HARX models coefficients.

Coefficients	Measures	S&P 500	DJIA	RUT
β_0	Coefficient	0.0004	0.0005	0.0005
	T-statistics *p*-value	2.70 7.07E^{-3}***	2.92 3.65E^{-3}***	2.52 0.01**
2DWD	Coefficient	0.011	0.012	0.019
	T-statistics *p*-value	2.54 0.01**	2.69 7.44E^{-3}***	3.48 5.25E^{-4}***
$\overline{RV}_{t-1:t-22}$	Coefficient	0.329	0.349	0.336
	T-statistics *p*-value	8.06 3.23E^{-15}***	8.14 2.27E^{-15}***	8.08 2.90E^{-15}***
$\overline{RV}_{t-1:t-5}$	Coefficient	0.375	0.373	0.404
	T-statistics *p*-value	8.41 2.12E^{-16}***	8.09 3.25E^{-15}***	8.89 5.61E^{-18}***
RV_{t-1}	Coefficient	0.025	0.025	0.138
	T-statistics *p*-value	7.21 1.36E^{-12}***	5.91 5.78E^{-9}***	5.46 6.81E^{-8}***

Table 8
HARX models coefficients with control variables.

Coefficients	Measures	S&P 500	DJIA	RUT
β_0	Coefficient	−0.002	−0.001	0.0002
	T-statistics *p*-value	−7.53 2.55E^{-13}***	−5.47 8.00E^{-8}***	1.35 0.17
2DWD	Coefficient	0.015	0.014	0.016
	T-statistics *p*-value	3.70 2.45E^{-4}***	3.35 8.70E^{-4}***	3.23 1.34E^{-3}***
$\overline{RV}_{t-1:t-22}$	Coefficient	−0.170	−0.108	−0.056
	T-statistics *p*-value	−3.36 8.37E^{-4}***	−2.06 0.04**	−1.14 0.25
$\overline{RV}_{t-1:t-5}$	Coefficient	0.188	0.213	0.272
	T-statistics *p*-value	4.22 2.96E^{-5}***	4.63 4.92E^{-6}***	6.04 3.34E^{-9}***
RV_{t-1}	Coefficient	0.058	0.048	0.037
	T-statistics *p*-value	2.46 0.01**	2.03 0.04**	1.61 0.10
VIX	Coefficient	0.0004	0.0003	0.0004
	T-statistics *p*-value	14.65 1.63E^{-40}***	13.18 3.38E^{-33}***	12.81 3.70E^{-32}***
Mkt-RF	Coefficient	−0.0005	−0.0005	−0.0008
	T-statistics *p*-value	−6.88 1.86E^{-11}***	−7.16 3.83E^{-12}***	−8.24 1.99E^{-15}***
SMB	Coefficient	−0.0002	−0.0001	−0.0001
	T-statistics *p*-value	−1.25 0.02**	−0.93 0.35	−0.90 0.37
HML	Coefficient	0.0002	0.0001	0.0003
	T-statistics *p*-value	1.42 0.15	−1.07 0.28	1.77 0.07*

Table 9
Full sample results.

Stock Index	Measures	NBEATS	NBEATSx-PH	HAR	HARX	HARST-Week	HARST-PH
	RMSE	0.475%	**0.416%**	0.467%	0.461%	0.467%	0.452%
	MAE	0.293%	**0.269%**	0.291%	0.292%	0.290%	0.285%
S&P 500	QLIKE	9.40%	**8.65%**	10.03%	9.67%	10.06%	9.67%
	DM test (RMSE)	3.31E^{-3}***	3.31E^{-3}***	0.03**	0.03**	4.13E^{-3}***	4.13E^{-3}***
	DM test (MAE)	7.33E^{-5}***	7.33E^{-5}***	0.78	0.78	2.95E^{-3}***	2.95E^{-3}***
	DM test (QLIKE)	1.53E^{-3}***	1.53E^{-3}***	0.02**	0.02**	9.85E^{-5}***	9.85E^{-5}***
	MCS test (RMSE)	0.02**	**1.00**	0.02**	0.02**	0.02**	0.02**
	MCS test (MAE)	5.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (QLIKE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	1.00E^{-3}***	>1.00E^{-3}***	2.00E^{-3}***
	RMSE	0.542%	**0.451%**	0.540%	0.533%	0.543%	0.536%
	MAE	0.306%	**0.266%**	0.306%	0.305%	0.304%	0.306%
DJIA	QLIKE	10.96%	**8.18%**	10.05%	9.77%	10.16%	9.87%
	DM test (RMSE)	0.02**	0.02**	0.02**	0.02**	0.20	0.20
	DM test (MAE)	7.07E^{-8}***	7.07E^{-8}***	0.63	0.63	0.46	0.46
	DM test (QLIKE)	1.40E^{-12}***	1.40E^{-12}***	4.96E^{-3}***	4.96E^{-3}***	9.16E^{-4}***	9.16E^{-4}***
	MCS test (RMSE)	0.05*	**1.00**	0.05*	0.05*	0.05*	0.05*
	MCS test (MAE)	0.02**	**1.00**	0.02**	0.02**	0.02**	0.02**
	MCS test (QLIKE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	RMSE	0.483%	**0.439%**	0.500%	0.494%	0.495%	0.484%
	MAE	0.320%	**0.305%**	0.344%	0.345%	0.342%	0.341%
RUT	QLIKE	9.26%	**8.23%**	9.17%	9.03%	9.14%	8.93%
	DM test (RMSE)	1.84E^{-3}***	1.84E^{-3}***	0.16	0.16	0.07*	0.07*
	DM test (MAE)	0.01**	0.01**	0.65	0.65	0.50	0.50
	DM test (QLIKE)	1.70E^{-3}***	1.70E^{-3}***	0.14	0.14	2.97E^{-3}***	2.97E^{-3}***
	MCS test (RMSE)	0.02**	**1.00**	0.02**	0.02**	0.02**	0.02**
	MCS test (MAE)	0.01**	**1.00**	1.00E^{-3}***	1.00E^{-3}***	1.00E^{-3}***	1.00E^{-3}***
	MCS test (QLIKE)	1.00E^{-3}***	**1.00**	0.02**	0.06*	0.02**	0.06*

The *p*-values shown for the DM test are only concerning the pair of models in the same model family (i.e., NBEATS vs NBEATSx-PH, HAR vs HARX, and HARST-week vs HARST-PH). Additionally, the best values out of all models are in bold to facilitate the read of the table.

Table 10
Results for the 2020 crisis.

Stock Index	Measures	NBEATS	NBEATSx-PH	HAR	HARX	HARST-Week	HARST-PH
	RMSE	1.284%	**0.953%**	1.226%	1.186%	1.228%	1.163%
	MAE	0.907%	**0.677%**	0.887%	0.856%	0.881%	0.842%
S&P 500	QLIKE	10.32%	**5.80%**	11.50%	8.79%	11.67%	9.96%
	DM test (RMSE)	0.01**	0.01**	0.03**	0.03**	0.03**	0.03**
	DM test (MAE)	5.27E^{-3}***	5.27E^{-3}***	0.08*	0.08*	0.04**	0.04**
	DM test (QLIKE)	7.81E^{-4}***	7.81E^{-4}***	0.05*	0.05*	2.03E^{-3}***	2.03E^{-3}***
	MCS test (RMSE)	1.00E^{-3}***	**1.00**	1.00E^{-3}***	1.00E^{-3}***	1.00E^{-3}***	1.00E^{-3}***
	MCS test (MAE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (QLIKE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	RMSE	1.544%	**1.177%**	1.587%	1.550%	1.607%	1.569%
	MAE	0.982%	**0.725%**	1.050%	1.010%	1.048%	1.044%
DJIA	QLIKE	14.13%	**6.05%**	13.51%	10.72%	14.00%	12.42%
	DM test (RMSE)	0.12	0.12	0.03**	0.03**	0.27	0.27
	DM test (MAE)	3.76E^{-3}***	3.76E^{-3}***	0.02**	0.02**	0.88	0.88
	DM test (QLIKE)	2.19E^{-4}***	2.19E^{-4}***	0.02**	0.02**	0.02**	0.02**
	MCS test (RMSE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (MAE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (QLIKE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	RMSE	1.048%	**0.773%**	1.093%	1.042%	1.063%	0.984%
	MAE	0.731%	**0.580%**	0.850%	0.809%	0.820%	0.785%
RUT	QLIKE	7.62%	**4.85%**	9.49%	6.80%	8.99%	6.86%
	DM test (RMSE)	0.01**	0.01**	0.09*	0.09*	0.10	0.10
	DM test (MAE)	0.03**	0.03**	0.19	0.19	0.34	0.34
	DM test (QLIKE)	7.25E^{-3}***	7.25E^{-3}***	0.01**	0.01**	4.85E^{-3}***	4.85E^{-3}***
	MCS test (RMSE)	3.00E^{-3}***	**1.00**	>1.00E^{-3}***	3.00E^{-3}***	3.00E^{-3}***	>1.00E^{-3}***
	MCS test (MAE)	4.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (QLIKE)	2.00E^{-3}***	**1.00**	>1.00E^{-3}***	2.00E^{-3}***	1.00E^{-3}***	>1.00E^{-3}***

The *p*-values shown for the DM test are only concerning the pair of models in the same model family (i.e., NBEATS vs NBEATSx-PH, HAR vs HARX, and HARST-week vs HARST-PH). Additionally, the best values out of all models are in bold to facilitate the read of the table.

Table 11
HARX models coefficients.

Coefficients	Measures	S&P 500	DJIA	RUT
β_0	Coefficient	0.0005	0.0006	0.0008
	T-statistics *p*-value	3.50 4.86E^{-4}***	3.72 2.07E^{-4}***	4.14 3.88E^{-5}***
2DWD	Coefficient	0.019	0.017	0.015
	T-statistics *p*-value	6.40 2.36E^{-10}***	6.01 2.66E^{-9}***	4.80 1.90E^{-6}***
$\overline{RV}_{t-1:t-22}$	Coefficient	0.222	0.206	0.245
	T-statistics *p*-value	6.53 1.07E^{-10}***	6.02 2.53E^{-9}***	7.04 4.03E^{-12}***
$\overline{RV}_{t-1:t-5}$	Coefficient	0.519	0.560	0.474
	T-statistics *p*-value	13.80 7.42E^{-40}***	14.86 3.59E^{-45}***	12.62 1.55E^{-33}***
RV_{t-1}	Coefficient	0.112	0.086	0.146
	T-statistics *p*-value	5.34 1.16E^{-7}***	4.12 4.15E^{-5}***	6.98 6.17E^{-12}***

Table 12
HARX models coefficients with control variables.

Coefficients	Measures	S&P 500	DJIA	RUT
β_0	Coefficient	−0.001	−0.0007	0.0002
	T-statistics *p*-value	−5.58 3.57E^{-8}***	−3.59 3.65E^{-4}***	1.09 0.27
2DWD	Coefficient	0.019	0.016	0.006
	T-statistics *p*-value	6.69 4.91E^{-11}***	5.65 2.46E^{-8}***	2.00 0.04**
$\overline{RV}_{t-1:t-22}$	Coefficient	−0.119	−0.072	0.133
	T-statistics *p*-value	−2.83 4.80E^{-3}***	−1.75 0.08*	3.62 3.28E^{-4}***
$\overline{RV}_{t-1:t-5}$	Coefficient	0.426	0.489	0.461
	T-statistics *p*-value	11.60 2.32E^{-28}***	13.30 1.73E^{-35}***	12.74 1.96E^{-32}***
RV_{t-1}	Coefficient	0.036	0.023	0.095
	T-statistics *p*-value	1.80 0.07*	1.14 0.25	4.74 2.79E^{-6}***
VIX	Coefficient	0.0003	0.0002	0.0001
	T-statistics *p*-value	12.66 6.11E^{-33}***	11.15 2.43E^{-26}***	7.46 3.69E^{-13}***
Mkt-RF	Coefficient	−0.0006	−0.0006	−0.0008
	T-statistics *p*-value	−11.16 1.44E^{-26}***	−11.07 5.28E^{-26}***	−13.62 3.41E^{-36}***
SMB	Coefficient	−0.0005	−0.0005	−0.0008
	T-statistics *p*-value	−4.99 7.66E^{-7}***	−4.75 2.61E^{-6}***	−6.55 1.45E^{-10}***
HML	Coefficient	4.34E^{-5}	−0.0001	0.0001
	T-statistics *p*-value	0.45 0.65	−1.16 0.24	1.06 0.28

Table 13
Full sample results.

Stock Index	Measures	NBEATS	NBEATSx-PH	HAR	HARX	HARST-Week	HARST-PH
	RMSE	0.358%	**0.340%**	0.374%	0.371%	0.372%	0.370%
	MAE	0.226%	**0.224%**	0.239%	0.240%	0.240%	0.240%
S&P 500	QLIKE	8.70%	**8.63%**	9.46%	9.40%	9.57%	9.65%
	DM test (RMSE)	$2.47E^{-11}$***	$2.47E^{-11}$***	0.05*	0.05*	0.09*	0.09*
	DM test (MAE)	$8.97E^{-18}$***	$8.97E^{-18}$***	0.41	0.41	$8.13E^{-4}$***	$8.13E^{-4}$***
	DM test (QLIKE)	$3.01E^{-6}$***	$3.01E^{-6}$***	0.07*	0.07*	0.08*	0.08*
	MCS test (RMSE)	0.26	**1.00**	0.09*	0.09*	0.09*	0.04**
	MCS test (MAE)	0.47	**1.00**	$1.00E^{-3}$**	$1.00E^{-3}$***	$>1.00E^{-3}$***	$>1.00E^{-3}$***
	MCS test (QLIKE)	0.58	**1.00**	0.02**	0.02**	$>1.00E^{-3}$***	$>1.00E^{-3}$***
	RMSE	0.419%	**0.376%**	0.418%	0.414%	0.414%	0.408%
	MAE	0.234%	**0.232%**	0.244%	0.245%	0.249%	0.244%
DJIA	QLIKE	9.58%	10.12%	9.42%	**9.33%**	9.52%	9.45%
	DM test (RMSE)	0.13	0.13	0.07*	0.07*	0.28	0.28
	DM test (MAE)	0.49	0.49	0.57	0.57	$1.40E^{-3}$***	$1.40E^{-3}$***
	DM test (QLIKE)	0.03**	0.03**	0.19	0.19	0.40	0.40
	MCS test (RMSE)	0.47	**1.00**	0.42	0.47	0.47	0.47
	MCS test (MAE)	0.58	**1.00**	0.06*	0.06*	$>1.00E^{-3}$***	0.04**
	MCS test (QLIKE)	0.48	0.48	0.31	**1.00**	$>1.00E^{-3}$***	0.02**
	RMSE	0.440%	**0.398%**	0.418%	0.414%	0.413%	0.407%
	MAE	0.309%	**0.278%**	0.292%	0.291%	0.294%	0.289%
RUT	QLIKE	10.11%	9.25%	8.66%	**8.58%**	8.66%	8.56%
	DM test (RMSE)	$2.47E^{-11}$***	$2.47E^{-11}$***	0.05*	0.05*	0.09*	0.09*
	DM test (MAE)	$8.97E^{-18}$***	$8.97E^{-18}$***	0.41	0.41	$8.13E^{-4}$***	$8.13E^{-4}$***
	DM test (QLIKE)	$3.01E^{-6}$***	$3.01E^{-6}$***	0.07*	0.07*	0.08*	0.08*
	MCS test (RMSE)	$3.00E^{-3}$***	**1.00**	0.20	0.40	0.40	0.56
	MCS test (MAE)	$6.00E^{-3}$***	**1.00**	0.27	0.45	$7.00E^{-3}$***	0.45
	MCS test (QLIKE)	$2.00E^{-3}$***	0.52	0.05*	0.52	0.27	**1.00**

The *p*-values shown for the DM test are only concerning the pair of models in the same model family (i.e., NBEATS vs NBEATSx-PH, HAR vs HARX, and HARST-week vs HARST-PH). Additionally, the best values out of all models are in bold to facilitate the read of the table.

Table 14
Results for the 2020 crisis.

Stock Index	Measures	NBEATS	NBEATSx-PH	HAR	HARX	HARST-Week	HARST-PH
	RMSE	1.078%	**0.894%**	1.186%	1.140%	1.161%	1.092%
	MAE	0.751%	**0.631%**	0.886%	0.822%	0.849%	0.814%
S&P 500	QLIKE	6.86%	**4.82%**	10.71%	7.57%	10.16%	8.02%
	DM test (RMSE)	0.07*	0.07*	0.06*	0.06*	0.18	0.18
	DM test (MAE)	0.03**	0.03**	0.10	0.10	0.10	0.10
	DM test (QLIKE)	0.01**	0.01**	0.08*	0.08*	0.02**	0.02**
	MCS test (RMSE)	$2.00E^{-3}$***	**1.00**	$2.00E^{-3}$***	$2.00E^{-3}$***	$2.00E^{-3}$***	$2.00E^{-3}$***
	MCS test (MAE)	$>1.00E^{-3}$***	**1.00**	$>1.00E^{-3}$***	$>1.00E^{-3}$***	$>1.00E^{-3}$***	$>1.00E^{-3}$***
	MCS test (QLIKE)	$>1.00E^{-3}$***	**1.00**	$>1.00E^{-3}$***	$>1.00E^{-3}$***	$>1.00E^{-3}$***	$>1.00E^{-3}$***
	RMSE	1.526%	**1.074%**	1.517%	1.477%	1.479%	1.398%
	MAE	0.916%	**0.687%**	1.009%	0.966%	0.970%	0.925%
DJIA	QLIKE	12.43%	**5.46%**	11.87%	9.05%	10.83%	8.87%
	DM test (RMSE)	0.11	0.11	0.03**	0.03**	0.20	0.20
	DM test (MAE)	0.02**	0.02**	0.07*	0.07*	0.19	0.19
	DM test (QLIKE)	$2.31E^{-3}$***	$2.31E^{-3}$***	0.05*	0.05*	0.04**	0.04**
	MCS test (RMSE)	$7.00E^{-3}$***	**1.00**	$>1.00E^{-3}$***	$7.00E^{-3}$***	$7.00E^{-3}$***	$7.00E^{-3}$***
	MCS test (MAE)	$>1.00E^{-3}$***	**1.00**	$>1.00E^{-3}$***	$>1.00E^{-3}$***	$>1.00E^{-3}$***	$>1.00E^{-3}$***
	MCS test (QLIKE)	$>1.00E^{-3}$***	**1.00**	$>1.00E^{-3}$***	$>1.00E^{-3}$***	$>1.00E^{-3}$***	$>1.00E^{-3}$***
	RMSE	0.918%	**0.698%**	1.038%	0.955%	0.976%	0.875%
	MAE	0.697%	**0.523%**	0.808%	0.772%	0.750%	0.704%
RUT	QLIKE	6.42%	**3.85%**	8.55%	6.42%	7.92%	5.67%
	DM test (RMSE)	$4.16E^{-4}$***	$4.16E^{-4}$***	0.04**	0.04**	0.02**	0.02**
	DM test (MAE)	$2.64E^{-4}$***	$2.64E^{-4}$***	0.13	0.13	0.09*	0.09*
	DM test (QLIKE)	$9.57E^{-5}$***	$9.57E^{-5}$***	0.01**	0.01**	0.02**	0.02**
	MCS test (RMSE)	$>1.00E^{-3}$***	**1.00**	$>1.00E^{-3}$***	$>1.00E^{-3}$***	$>1.00E^{-3}$***	$>1.00E^{-3}$***
	MCS test (MAE)	$>1.00E^{-3}$***	**1.00**	$>1.00E^{-3}$***	$>1.00E^{-3}$***	$>1.00E^{-3}$***	$>1.00E^{-3}$***
	MCS test (QLIKE)	$>1.00E^{-3}$***	**1.00**	$>1.00E^{-3}$***	$>1.00E^{-3}$***	$>1.00E^{-3}$***	$>1.00E^{-3}$***

The *p*-values shown for the DM test are only concerning the pair of models in the same model family (i.e., NBEATS vs NBEATSx-PH, HAR vs HARX, and HARST-week vs HARST-PH). Additionally, the best values out of all models are in bold to facilitate the read of the table.

Table 15
HARX models coefficients.

Coefficients	Measures	STOXX	GDAXI	FTSE
β_0	Coefficient	0.0005	0.0007	0.0005
	T-statistics *p*-value	3.63 3.02E^{-4}***	3.88 1.14E^{-4}***	3.42 6.57E^{-4}***
2DWD	Coefficient	0.019	0.020	0.018
	T-statistics *p*-value	6.29 4.92E^{-10}***	5.55 3.75E^{-8}***	6.24 6.28E^{-10}***
$\overline{RV}_{t-1:t-22}$	Coefficient	0.295	0.266	0.301
	T-statistics *p*-value	8.61 3.44E^{-17}***	8.11 1.65E^{-15}***	9.02 8.94E^{-19}***
$\overline{RV}_{t-1:t-5}$	Coefficient	0.345	0.387	0.353
	T-statistics *p*-value	9.53 1.57E^{-20}***	10.86 6.06E^{-26}***	9.73 1.73E^{-21}***
RV_{t-1}	Coefficient	0.205	0.208	0.206
	T-statistics *p*-value	10.23 2.91E^{-23}***	10.33 9.74E^{-24}***	10.23 1.87E^{-23}***

Table 16
HARX models coefficients with control variables.

Coefficients	Measures	STOXX	GDAXI	FTSE
β_0	Coefficient	−0.0001	0.0002	−0.0003
	T-statistics *p*-value	−0.74 0.45	0.91 0.36	1.98 0.04**
2DWD	Coefficient	0.015	0.010	0.018
	T-statistics *p*-value	5.29 1.78E^{-7}***	2.90 3.85E^{-3}***	6.64 6.39E^{-11}***
$\overline{RV}_{t-1:t-22}$	Coefficient	−0.0133	0.064	−0.072
	T-statistics *p*-value	−0.33 0.74	1.76 0.07*	−1.75 0.07*
$\overline{RV}_{t-1:t-5}$	Coefficient	0.275	0.340	0.267
	T-statistics *p*-value	7.70 6.06E^{-14}***	9.69 1.05E^{-20}***	7.37 4.74E^{-13}***
RV_{t-1}	Coefficient	0.146	0.163	0.141
	T-statistics *p*-value	7.46 3.30E^{-13}***	8.23 1.21E^{-15}***	7.14 2.41E^{-12}***
VIX	Coefficient	0.0002	0.0002	0.0003
	T-statistics *p*-value	13.14 1.14E^{-34}***	11.61 3.16E^{-28}***	14.37 3.92E^{-41}***
Mkt-RF	Coefficient	−0.0004	−0.0004	−0.0002
	T-statistics *p*-value	−6.54 1.40E^{-10}***	−5.37 1.13E^{-7}***	−4.50 8.09E^{-6}***
SMB	Coefficient	−0.0002	−0.0005	−0.0002
	T-statistics *p*-value	−1.78 0.07*	−3.76 1.84E^{-4}***	−1.54 0.12

for at least one error measure, with the exception of HAR and HARX models for S&P 500 when considering the *p*-value threshold of 0.05. Thus, it can be concluded that the consequences of the Topological Tail Dependence Theory indeed agree with main American stock indexes for at least the 21st century. Incidentally, NBEATSx-PH is anew the best model among all other models during the 2020 crisis.

6.3. *Robustness test (iii)*

Table 15 presents the coefficients of the HARX models for robustness test (iii). Similar to the American stock indexes, the coefficients for 2DWD are all statistically significant at the level 0.01, further demonstrating that 2DWD likely has a significant linear relationship with RV. Incidentally, it is worth mentioning that the 2DWD here is the 2DWD estimated by the PH-RV algorithm with the data of the considered European stock indexes.

The estimated coefficients of the HARX models with the incorporation of control variables can be found in Table 16. Although at first glance the use of VIX close prices and Fama-French's three factors as control variables for European stock indexes might seem a bit odd, when performing a quantitative exploration, their coefficients are all, apart from HML, statistically significant at the level of 0.01 for at least one stock index. Additionally, this can also be conceptually explained by 1. The high interconnectedness between the European financial markets and the American financial market due to the free and fast flow of information and capital (Baele et al., 2004), 2. The universality of Fama-French's three factors (Asness et al., 2013; Fama and French, 2015), 3. Digitalization has facilitated the rapid dissemination of financial information globally, leading American real-time data and news to flow seamlessly across borders, impacting investor behavior and market sentiment in Europe (Baele et al., 2004). However, it is noteworthy that HML was not added as a control variable for the European stock indexes since it was not statistically significant at the level of 0.10 for any stock index.

Similar to the American stock indexes, the coefficients for 2DWD are all statistically significant at the level of 0.01 for all stock indexes; thus, confirming that there exists a statistically significant linear relationship between 2DWD and RV of American and European stock indexes.

Table 17, on the other hand, presents the error measures and *p*-values of the statistical tests for the full test sample of the European stock indexes. Anew, NBEATSx-PH is superior to all other models, with the exception when considering QLIKE for GDAXI.

Interestingly, although the models with PH information systematically have better forecast performance for all stock indexes when considering RMSE and QLIKE, the added value of the incorporation of PH information for forecast accuracy is less clear than for the American stock indexes. Only when considering RMSE for NBEATS vs NBEATSx-PH, this added value is statistically significant at the level of 0.05. However, this is not enough to falsify the Topological Tail Dependence Theory since it predicts that the incorporation of PH

Table 17
Full sample results.

Stock Index	Measures	NBEATS	NBEATSx-PH	HAR	HARX	HARST-Week	HARST-PH
	RMSE	0.356%	**0.345%**	0.415%	0.409%	0.412%	0.395%
	MAE	0.231%	**0.218%**	0.260%	0.265%	0.261%	0.263%
STOXX	QLIKE	7.96%	**7.70%**	9.95%	9.90%	9.93%	9.89%
	DM test (RMSE)	0.04**	0.04**	0.39	0.39	0.07*	0.07*
	DM test (MAE)	0.65	0.65	0.46	0.46	0.10	0.10
	DM test (QLIKE)	0.12	0.12	0.66	0.66	0.02**	0.02**
	MCS test (RMSE)	0.02**	**1.00**	0.02**	0.02**	0.02**	0.01**
	MCS test (MAE)	0.23	**1.00**	>1.00E^{-3}**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (QLIKE)	0.01**	**1.00**	0.01**	2.00E^{-3}***	0.01**	0.01**
	RMSE	0.523%	**0.489%**	0.595%	0.588%	0.591%	0.569%
	MAE	0.332%	**0.312%**	0.377%	0.380%	0.376%	0.376%
GDAXI	QLIKE	**8.66%**	8.93%	10.48%	10.34%	10.43%	10.12%
	DM test (RMSE)	0.02**	0.02**	0.41	0.41	0.19	0.19
	DM test (MAE)	4.68E^{-3}***	4.68E^{-3}***	0.22	0.22	0.93	0.93
	DM test (QLIKE)	0.41	0.41	0.37	0.37	0.09*	0.09*
	MCS test (RMSE)	0.14	**1.00**	0.04**	0.04**	0.04**	0.02**
	MCS test (MAE)	0.04**	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (QLIKE)	0.24	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	RMSE	0.371%	**0.356%**	0.408%	0.404%	0.415%	0.395%
	MAE	0.240%	**0.237%**	0.266%	0.267%	0.267%	0.262%
FTSE	QLIKE	7.39%	**6.93%**	8.24%	8.18%	8.30%	8.01%
	DM test (RMSE)	0.04**	0.04**	0.39	0.39	0.07*	0.07*
	DM test (MAE)	0.65	0.65	0.46	0.46	0.10	0.10
	DM test (QLIKE)	0.12	0.12	0.66	0.66	0.02**	0.02**
	MCS test (RMSE)	0.02**	**1.00**	0.02**	0.02**	0.02**	0.02**
	MCS test (MAE)	0.25	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (QLIKE)	0.01**	**1.00**	0.01**	5.00E^{-3}***	0.01**	0.01**

The *p*-values shown for the DM test are only concerning the pair of models in the same model family (i.e., NBEATS vs NBEATSx-PH, HAR vs HARX, and HARST-week vs HARST-PH). Additionally, the best values out of all models are in bold to facilitate the read of the table.

Table 18
Results for the 2020 crisis.

Stock Index	Measures	NBEATS	NBEATSx-PH	HAR	HARX	HARST-Week	HARST-PH
	RMSE	0.791%	**0.751%**	1.034%	0.994%	1.014%	0.900%
	MAE	0.566%	**0.526%**	0.703%	0.693%	0.691%	0.660%
STOXX	QLIKE	8.99%	**5.68%**	13.10%	10.11%	12.88%	9.62%
	DM test (RMSE)	0.29	0.29	0.17	0.17	0.14	0.14
	DM test (MAE)	0.28	0.28	0.73	0.73	0.54	0.54
	DM test (QLIKE)	0.04**	0.04**	0.08*	0.08*	0.04**	0.04**
	MCS test (RMSE)	1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (MAE)	4.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	1.00E^{-3}***
	MCS test (QLIKE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	RMSE	1.271%	**1.061%**	1.537%	1.500%	1.515%	1.385%
	MAE	0.957%	**0.760%**	1.112%	1.109%	1.087%	1.070%
GDAXI	QLIKE	8.56%	**7.61%**	14.71%	11.68%	14.26%	9.41%
	DM test (RMSE)	0.01**	0.01**	0.37	0.37	0.18	0.18
	DM test (MAE)	7.58E^{-3}***	7.58E^{-3}***	0.92	0.92	0.77	0.77
	DM test (QLIKE)	0.44	0.44	0.09*	0.09*	0.04**	0.04**
	MCS test (RMSE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***
	MCS test (MAE)	1.00E^{-3}***	**1.00**	>1.00E^{-3}***	1.00E^{-3}***	>1.00E^{-3}***	1.00E^{-3}***
	MCS test (QLIKE)	0.01**	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	4.00E^{-3}***
	RMSE	0.889%	**0.826%**	1.036%	1.001%	1.070%	0.965%
	MAE	0.672%	**0.613%**	0.755%	0.736%	0.786%	0.714%
FTSE	QLIKE	9.02%	**6.70%**	12.79%	10.25%	13.37%	10.24%
	DM test (RMSE)	0.05**	0.05**	0.25	0.25	0.08*	0.08*
	DM test (MAE)	0.10	0.10	0.48	0.48	0.09*	0.09*
	DM test (QLIKE)	0.02**	0.02**	0.10	0.10	0.01**	0.01**
	MCS test (RMSE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	1.00E^{-3}***
	MCS test (MAE)	3.00E^{-3}***	**1.00**	>1.00E^{-3}***	0.01**	>1.00E^{-3}***	0.01**
	MCS test (QLIKE)	>1.00E^{-3}***	**1.00**	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***	>1.00E^{-3}***

The *p*-values shown for the DM test are only concerning the pair of models in the same model family (i.e., NBEATS vs NBEATSx-PH, HAR vs HARX, and HARST-week vs HARST-PH). Additionally, the best values out of all models are in bold to facilitate the read of the table.

information ought to improve the model's forecast performance during *turbulent periods*. Hence, the forecasting results of the models during the 2020 crisis must be analysed.

Table 18 shows the forecasting results for the 2020 crisis. Again, as predicted by the Topological Tail Dependence Theory, the models that incorporate PH information systematically yield statistically more accurate forecasts for at least one error measure, apart from the linear models at the level of 0.05. Thus, it can be concluded that the employment of PH information allows nonlinear and neural network models to better forecast RV during turbulent period, yet the same cannot be said with certainty about linear models.

In conclusion, it can be affirmed that the results of the main sample and robustness tests fails to falsify the Topological Tail Dependence Theory; thus, increasing the likelihood that it is a correct theory that explains the financial reason behind the recent success of the employment of PH in foreseeing financially turbulent periods and improving portfolio management. Additionally, the results of the main sample and robustness tests demonstrate that the incorporation of PH information into RV forecasting models can be beneficial, especially during turbulent periods.

7. Conclusion

This paper proposes a novel theory, coined Topological Tail Dependence Theory, that links the mathematical theory behind Persistent Homology (PH) and the financial stock market theory. This study also proposes a novel algorithm to measure topological stock market changes as well as the incorporation of these topological changes into forecasting realized volatility (RV) models to improve their forecast performance during turbulent periods.

In broad terms, the Topological Tail Dependence Theory claims Wasserstein Distances (WD) or L^n norms of Persistent Landscape, estimated through PH information and used as a proxy for stock market change, can be used as an instrument to foresee financially turbulent periods by exploring the fact that correlations between stocks increase before and during turbulent periods. The use of these topological measures over simple correlation metrics is motivated by the fact that these measures do not suffer from the curse of dimensionality and can detect nonlinear and complex correlations. One of the implications of this theory is that if PH information is used to forecast realized volatility, it should in any case improve the model's forecast performance during turbulent periods. This is the case since this additional information will allow the model to better foresee high values of stock realized volatility, which occur during turbulent periods.

As a result, the empirical part of this study is dedicated to testing this expectation by putting the proposed theory to the test with three different families of models, linear models, nonlinear models, and neural network models. The linear models are Heterogeneous Autoregressive (HAR) model (Corsi, 2009) and Heterogeneous Autoregressive with eXogenous variables (HARX) model (Hillebrand and Medeiros, 2010), while the nonlinear models are Multiple-Regime Smooth Transition Heterogeneous Autoregressive (HARST) model (McAleer and Medeiros, 2008) with the weekly moving average of RV and 2d Wasserstein Distances (2DWD) as transition variables. Lastly, the neural network models are Neural Basis Expansion Analysis with incorporate Exogenous factors (NBEATSx).

This research tests the proposed theory by exploring three famous stock indexes, Standard & Poor's 500 (S&P 500), Dow Jones Industrial Average (DJIA) and Russell 2000 (RUT), from 1st of January 2000 until 30 March 2022, with a data splitting of 80% training data and 20% testing data. To test the results found with the main sample, this paper also performs two robustness tests with the same sample (training size reduced by a half (i) and different data splitting 60%/40% (ii)) and explore stock indexes from different geographical locations, Financial Times Stock Exchange 100 (FTSE), DAX Performance (GDAXI), and STXE 600 (STOXX), as a third robustness test (iii). Additionally, this study employs Root-Mean-Square Error (RMSE), Mean Absolute Error (MAE), and Quasi-likelihood (QLIKE) as error measures, and performs Diebold-Mariano (DM) tests and Model Confidence Set (MCS) tests for each error measure.

The findings of the main sample show that the coefficients of 2DWD are statistically significant for the linear models even when accounting for control variables. These results are equal for all robustness tests. Consequently, it can be concluded that there exists a statistically significant linear and positive relationship between 2DWD and future RV. Concerning the forecasts for the full testing sample, the incorporation of PH information into the models systematically improves the model's forecast accuracy (and statistically significantly for at least one error measure, with the exception of the linear models for the second and third robustness tests). Regarding the forecasts during a turbulent period in the full testing sample, namely the 2020 crisis, all models with PH information systematically yield statistically significantly more precise forecasts than the other models for at least one measure for the main sample and the first two robustness tests. For the third robustness test, this is only true for the nonlinear and neural network models. This improvement in forecasting accuracy during turbulent periods is exactly what the Topological Tail Dependence Theory predicts. Hence, it can be concluded with certainty that the employment of PH information allows nonlinear and neural network models to better forecast RV during turbulent periods, yet the same cannot be said about linear models. Additionally, it can also be concluded that the empirical experimentation of this study fails to falsify the Topological Tail Dependence Theory, and thus this theory ought to be further explored by the scientific community.

Future research that can be done based on the findings of this paper is: 1. Further empirical experimentation of the Topological Tail Dependence Theory to further put it to the test, 2. Further develop the proposed theory and its logical implications in order to further test it, and 3. Explore the application of PH information into other financial domains.

Data and code availability

Data may be made available upon request. For code, see https://github.com/hugogobato/Topological-Tail-Dependence-Evidence-from-Forecasting-Realized-Volatility.git.

Declaration of competing 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.

Acknowledgments

The author of this paper would like to thank Jasmijn Merlijn Heuvel for her help with checking and improving the language use in this paper. Additionally, the author of this paper would like to thank Sérgio Paulo Amaral Souto and Yara Galvão Gobato for being great sources of inspiration for the author's academic career and personal life.

Last but not least, the author of this paper would like to thank the Editor-in-Chief as well as the two anonymous reviewers of this paper for giving highly valuable feedback that was essential for the improvement of this study. Finally, this study did not receive any funding of any kind.

Appendix A

The choice of using VIX close prices and Fama-French's three factors as control variables in linear regressions for forecasting realized volatility of S&P 500, DJIA, and RUT is based on the following logical and empirically supported reasons:

1. **VIX as a Fear Indicator:** VIX, often referred to as the "Fear Index," measures market sentiment and can be used to successfully forecast expected future realized volatility (Dutta and Das, 2022). It tends to rise during periods of uncertainty and market turbulence. Academic literature, such as the work by Whaley (2000), has shown that VIX is inversely related to stock market returns, while Dutta and Das (2022) have demonstrated that it can be used to forecast realized volatility. Therefore, including VIX close prices in regressions for stock index realized volatility can be logical. It serves as a leading indicator of potential turbulence and can help capture the influence of fear and market sentiment on realized volatility.
2. **Fama-French Factors and Risk Control:** Fama-French's three factors (market return, size, and value) are widely recognized in finance for explaining cross-sectional variations in stock returns. While they are often used to explain daily stock returns, they also have relevance in understanding the drivers of realized volatility (Liu and Zhang, 2021). Stocks in different size and value categories may exhibit varying levels of volatility, and these factors can help control for such systematic effects when modeling realized volatility.
3. **Link between Returns and Volatility:** There is a well-documented link between daily stock returns and realized volatility. Absolute higher returns tend to be associated with higher volatility. By including Fama-French factors that capture market returns, size, and value effects, the regressions account for some of the inherent relationships between returns and volatility, which can lead to more precise estimates of the determinants of realized volatility.

Thus, given the three arguments above, the choice of VIX close prices and Fama-French's three factors as control variables in the linear regressions for the HARX models comes naturally.

References

Albulescu, C.T., 2021. COVID-19 and the United States financial markets' volatility. Finance Res. Lett. 38, 101699. https://doi.org/10.1016/j.frl.2020.101699.
Alizadeh, S., Brandt, M.W., Diebold, F.X., 2002. Range-Based estimation of stochastic volatility models. J. Finance 57 (3), 1047–1091. https://doi.org/10.1111/1540-6261.00454.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71 (2), 579–625. https://doi.org/10.1111/1468-0262.00418.
Asness, C.S., Moskowitz, T.J., Pedersen, L.H., 2013. Value and momentum everywhere. J. Finance 68 (3), 929–985. https://doi.org/10.1111/jofi.12021.
Atkins, A., Niranjan, M., Gerding, E.H., 2018. Financial news predicts stock market volatility better than close price. The Journal of Finance and Data Science 4 (2), 120–137. https://doi.org/10.1016/j.jfds.2018.02.002.
Baele, L., Ferrando, A., Hördahl, P., Krylova, E., Monnet, C., 2004. Measuring financial integration in the euro area. RePEc: Research Papers in Economics 14 (1), 1–99. https://www.econstor.eu/bitstream/10419/154467/1/ecbop014.pdf.
Baitinger, E., Flegel, S., 2021. The better turbulence index? Forecasting adverse financial markets regimes with persistent homology. Financ. Mark. Portfolio Manag. 35 (3), 277–308. https://doi.org/10.1007/s11408-020-00377-x.
Beine, M., Cosma, A., Vermeulen, R., 2010. The dark side of global integration: increasing tail dependence. J. Bank. Finance 34 (1), 184–192. https://doi.org/10.1016/j.jbankfin.2009.07.014.
Bubenik, P., 2015. Statistical topological data analysis using persistence landscapes. J. Mach. Learn. Res. 16 (1), 77–102. https://dl.acm.org/doi/10.5555/2789272.2789275.
Buccheri, G., Corsi, F., 2019. HARK the SHARK: realized volatility modeling with measurement errors and nonlinear dependencies. J. Financ. Econom. 19 (4), 614–649. https://doi.org/10.1093/jjfinec/nbz025.
Carlsson, G.E., 2009. Topology and data. Bull. Am. Math. Soc. 46 (2), 255–308. https://doi.org/10.1090/s0273-0979-09-01249-x.
Chesnay, F., Jondeau, E., 2001. Does correlation between stock returns really increase during turbulent periods? Econ. Notes 30 (1), 53–80. https://doi.org/10.1111/1468-0300.00047.
Chiriac, R., Voev, V., 2010. Modelling and forecasting multivariate realized volatility. J. Appl. Econom. 26 (6), 922–947. https://doi.org/10.1002/jae.1152.
Cipollini, F., Gallo, G.M., Palandri, A., 2020. Realized variance modeling: decoupling forecasting from estimation. J. Financ. Econom. 18 (3), 532–555. https://doi.org/10.1093/jjfinec/nbaa009.
Corsi, F., 2009. A simple approximate Long-Memory model of realized volatility. J. Financ. Econom. 7 (2), 174–196. https://doi.org/10.1093/jjfinec/nbp001.
Corsi, F., Mittnik, S., Pigorsch, C., Pigorsch, U., 2008. The volatility of realized volatility. Econom. Rev. 27 (1–3), 46–78. https://doi.org/10.1080/07474930701853616.
Corsi, F., Pirino, D., Renò, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. J. Econom. 159 (2), 276–288. https://doi.org/10.1016/j.jeconom.2010.07.008.

Corsi, F., Audrino, F., Renò, R., 2012. HAR Modeling for Realized Volatility Forecasting. In: John Wiley & Sons, Inc. eBooks, pp. 363–382. https://doi.org/10.1002/9781118272039.ch15.
Deo, R.S., Hurvich, C.M., Lu, Y., 2006. Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment. J. Econom. 131 (1–2), 29–58. https://doi.org/10.1016/j.jeconom.2005.01.003.
Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econ. Stat. 13 (3), 253. https://doi.org/10.2307/1392185.
Dutta, A., Das, D., 2022. Forecasting realized volatility: New evidence from time-varying jumps in VIX. J. Futures Mark. 42 (12), 2165–2189. https://doi.org/10.1002/fut.22372.
Edelsbrunner, H., Harer, J., 2009. Computational Topology: an Introduction.
Elliott, G., Timmermann, A., 2008. Economic forecasting. J. Econ. Lit. 46 (1), 3–56. https://doi.org/10.1257/jel.46.1.3.
Engle, R.F., 2002. Dynamic conditional correlation. J. Bus. Econ. Stat. 20 (3), 339–350. https://doi.org/10.1198/073500102288618487.
Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. J. Financ. Econ. 116 (1), 1–22. https://doi.org/10.1016/j.jfineco.2014.10.010.
Gallo, G.M., Otranto, E., 2015. Forecasting realized volatility with changing average levels. Int. J. Forecast. 31 (3), 620–634. https://doi.org/10.1016/j.ijforecast.2014.09.005.
Gidea, M., 2017. Topological data analysis of critical transitions in financial networks. In: Springer Proceedings in Complexity. Springer International Publishing, pp. 47–59. https://doi.org/10.1007/978-3-319-55471-6_5.
Gidea, M., Katz, Y.A., 2018. Topological data analysis of financial time series: landscapes of crashes. Phys. Nonlinear Phenom. 491, 820–834. https://doi.org/10.1016/j.physa.2017.09.028.
Goel, A., Pasricha, P., Mehra, A., 2020. Topological data analysis in investment decisions. Expert Syst. Appl. 147, 113222. https://doi.org/10.1016/j.eswa.2020.113222.
Hansen, P.R., 2005. A test for superior predictive ability. J. Bus. Econ. Stat. 23 (4), 365–380. https://doi.org/10.1198/073500105000000063.
Hansen, P., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79 (2). https://doi.org/10.3982/ECTA5771.
Harvey, D., Leybourne, S.J., Newbold, P., 1997. Testing the equality of prediction mean squared errors. Int. J. Forecast. 13 (2), 281–291. https://doi.org/10.1016/s0169-2070(96)00719-4.
Hillebrand, E., Medeiros, M.C., 2010. The benefits of bagging for forecast models of realized volatility. Econom. Rev. 29 (5–6), 571–593. https://doi.org/10.1080/07474938.2010.481554.
Ismail, M.R., Noorani, M.S.M., Ismail, M., Razak, F.A., Alias, M.S., 2022. Early warning signals of financial crises using persistent homology. Phys. Stat. Mech. Appl. 586, 126459. https://doi.org/10.1016/j.physa.2021.126459.
Jebran, K., Chen, S., Ullah, I., Mirza, S.S., 2017. Does volatility spillover among stock markets varies from normal to turbulent periods? Evidence from emerging markets of Asia. The Journal of Finance and Data Science 3 (1–4), 20–30. https://doi.org/10.1016/j.jfds.2017.06.001.
Liu, H., Zhang, Q., 2021. Firm age and realized idiosyncratic return volatility in China: the role of short-sales constraints. Int. Rev. Financ. Anal. 75, 101745. https://doi.org/10.1016/j.irfa.2021.101745.
McAleer, M., Medeiros, M.C., 2008. A multiple regime smooth transition Heterogeneous Autoregressive model for long memory and asymmetries. J. Econom. 147 (1), 104–119. https://doi.org/10.1016/j.jeconom.2008.09.032.
McAleer, M., Medeiros, M.C., 2010. Forecasting realized volatility with linear and nonlinear univariate models. J. Econ. Surv. 25 (1), 6–18. https://doi.org/10.1111/j.1467-6419.2010.00640.x.
McInemey, T., Terzopoulos, D., 1999. Topology adaptive deformable surfaces for medical image volume segmentation. IEEE Trans. Med. Imag. 18 (10), 840–850. https://doi.org/10.1109/42.811261.
Moroni, D., Pascali, M.A., 2021. Learning topology: bridging computational topology and machine learning. Pattern Recogn. Image Anal. 31 (3), 443–453. https://doi.org/10.1134/s1054661821030184.
Nicolau, M., Levine, A.J., Carlsson, G.E., 2011. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. Proc. Natl. Acad. Sci. USA 108 (17), 7265–7270. https://doi.org/10.1073/pnas.1102826108.
Olivares, K.G., Challu, C., Marcjasz, G., Weron, R., Dubrawski, A., 2022. Neural basis expansion analysis with exogenous variables: forecasting electricity prices with NBEATSx. Int. J. Forecast. 39 (2), 884–900. https://doi.org/10.1016/j.ijforecast.2022.03.001.
Oreshkin, B.N., Dudek, G., Pełka, P., Turkina, E., 2021. N-BEATS neural network for mid-term electricity load forecasting. Appl. Energy 293, 116918. https://doi.org/10.1016/j.apenergy.2021.116918.
Pagan, A., Schwert, G.W., 1990. Alternative models for conditional stock volatility. J. Econom. 45 (1–2), 267–290. https://doi.org/10.1016/0304-4076(90)90101-x.
Patton, A.J., Sheppard, K., 2009. Evaluating Volatility and Correlation Forecasts. Springer eBooks, pp. 801–838. https://doi.org/10.1007/978-3-540-71297-8_36.
Patton, A.J., Sheppard, K., 2015. Good volatility, bad volatility: signed jumps and the persistence of volatility. Rev. Econ. Stat. 97 (3), 683–697. https://doi.org/10.1162/rest_a_00503.
Pereira, C.F., De Mello, R.F., 2015. Persistent homology for time series and spatial data clustering. Expert Syst. Appl. 42 (15–16), 6026–6038. https://doi.org/10.1016/j.eswa.2015.04.010.
Qiu, W., Rudkin, S., Dłotko, P., 2020. Refining understanding of corporate failure through a topological data analysis mapping of Altman's Z-score model. Expert Syst. Appl. 156, 113475. https://doi.org/10.1016/j.eswa.2020.113475.
Rogers, L.C.G., Satchell, S., 1991. Estimating variance from high, low and closing prices. Ann. Appl. Probab. 1 (4). https://doi.org/10.1214/aoap/1177005835.
Sbrana, A., Rossi, A.L.D., Naldi, M.C., 2020. N-BEATS-RNN: Deep Learning for Time Series Forecasting. 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA). https://doi.org/10.1109/icmla51294.2020.00125.
Shnier, D., Voineagu, M., Voineagu, I., 2019. Persistent homology analysis of brain transcriptome data in autism. J. R. Soc. Interface 16 (158), 20190531. https://doi.org/10.1098/rsif.2019.0531.
Souto, H.G., Moradi, A., 2023a. Introducing NBEATSx to realized volatility forecasting. Social Science Research Network (preprint). https://doi.org/10.2139/ssrn.4398498.
Souto, H.G., Moradi, A., 2023b. Realized covariance matrix NBEATSX. Social Science Research Network (preprint). https://doi.org/10.2139/ssrn.4529219.
Tralie, C.J., Saul, N., Bar-On, R., 2018. Ripser.py: a lean persistent homology library for Python. J. Open Source Softw. 3 (29), 925. https://doi.org/10.21105/joss.00925.
Tsantekidis, A., Passalis, N., Tefas, A., Kanniainen, J., Gabbouj, M., Iosifidis, A., 2017. Forecasting stock prices from the limit order book using convolutional neural networks. In: 2017 IEEE 19th Conference on Business Informatics (CBI). https://doi.org/10.1109/cbi.2017.23.
Wang, Y., Ma, F., Wei, Y., Wu, C., 2016. Forecasting realized volatility in a changing world: a dynamic model averaging approach. J. Bank. Finance 64, 136–149. https://doi.org/10.1016/j.jbankfin.2015.12.010.
Wang, X., Li, C., Yi, C., Xu, X., Wang, J., Zhang, Y., 2022. EcoForecast: an interpretable data-driven approach for short-term macroeconomic forecasting using N-BEATS neural network. Eng. Appl. Artif. Intell. 114, 105072. https://doi.org/10.1016/j.engappai.2022.105072.
Wasserman, L., 2017. Annual review of statistics and its application topological data analysis. Annual Reviews 5, 501–532. https://doi.org/10.1146/annurev-statistics-031017-10004.
Whaley, R.E., 2000. The investor fear gauge. J. Portfolio Manag. 26 (3), 12–17. https://doi.org/10.3905/jpm.2000.319728.
White, H., Kim, S.Y., Manganelli, S., 2015. VAR for VaR: measuring tail dependence using multivariate regression quantiles. J. Econom. 187 (1), 169–188. https://doi.org/10.1016/j.jeconom.2015.02.004.
Wong, Z.Y., Chin, W., Tan, S., 2016. Daily value-at-risk modeling and forecast evaluation: the realized volatility approach. The Journal of Finance and Data Science 2 (3), 171–187. https://doi.org/10.1016/j.jfds.2016.12.001.
Yang, D., Zhang, Q., 2000. Drift independent volatility estimation based on high, low, open, and close prices. J. Bus. 73 (3), 477–492. https://doi.org/10.1086/209650.
Yarovaya, L., Brzeszczyński, J., Lau, C.K.M., 2016. Intra- and inter-regional return and volatility spillovers across emerging and developed markets: evidence from stock indices and stock index futures. Int. Rev. Financ. Anal. 43, 96–114. https://doi.org/10.1016/j.irfa.2015.09.004.
Zhang, C., Pu, X., Cucuringu, M., Dong, X., 2023. Graph neural networks for forecasting realized volatility with nonlinear spillover effects. Social Science Research Network (preprint). https://doi.org/10.2139/ssrn.4375165.