RHINE: A Regime-Switching Model with Nonlinear Representation for Discovering and Forecasting Regimes in Financial Markets

Kunpeng Xu* Lifei Chen† Jean-Marc Patenaude ‡ Shengrui Wang*

Abstract

We investigate the problem of discovering and forecasting regular regime switches in a financial ecosystem comprising multiple time series. Such regime switches, indicative of varying market behaviors across distinct time intervals, are pivotal for a nuanced understanding of market dynamics, which in turn allows informed model selection for forecasting and enhanced interpretability of predictive outcomes. Despite strides in this domain, prevailing methodologies often falter due to: (1) an inability to effectively model the temporal behaviors inherent in financial series; and (2) neglecting the interdependencies among series when discovering regimes. In this paper, we propose RHINE, a **R**egime-switc**HI**ng model with **N**onlinear r**E**presentation. RHINE stands out with its kernel-based representation, adept at capturing the dynamic shifts in market regimes. This representation encapsulates the nonlinear interplay across multiple financial time series. By leveraging the kernel representation, we introduce an eigengap thresholding measure, designed to automatically discern the optimal number of financial market regimes, enhancing the model's adaptability to market fluctuations. Empirical assessments on both synthetic and real-world stock market datasets underscore RHINE's prowess. The findings illuminate that the inherent structures governing financial market behaviors are dynamic, and harnessing these dynamics via RHINE leads to a regime-based model that outperforms both conventional and state-of-the-art neural network models in predictive capabilities.

1 Introduction

Financial time series data, by its very nature, is dynamic and multifaceted, often exhibiting distinct behaviors across different time intervals. Such behaviors, or regimes, are not mere statistical anomalies but are reflective of underlying market dynamics. For instance, the six distinct segments depicted in Fig 1 showcase the market's cyclical nature. These segments are generally characterized by particular patterns in dynamically changing environments – *e.g.*, the volatility observed in stock returns during the global financial crisis of 2008-2009 was not a random fluctuation but a regime characterized by heightened uncertainty. Similarly, market reactions to geopolitical events, such as wars or peace talks, can be understood as regime switches. These regimes, be they transient like recessions or permanent like structural breaks, hold significant implications for investment strategies, risk management, and financial forecasting.

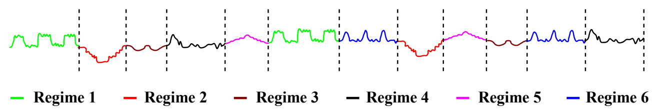


Figure 1. A time series involving six different noncontiguous regimes.

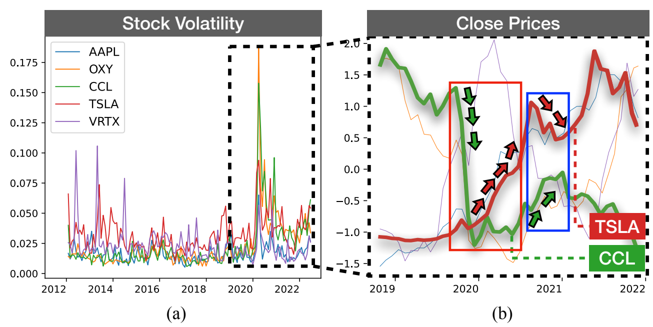


Figure 2. Time series of stocks. (a) Volatility of stocks in five different sectors in S&P 500. (b) Stock closing prices corresponding to volatility in 2019-2022.

To capture the dynamic market behaviors of time series, regime-switching (RS) models [21, 8] have been developed in the financial field. These studies suggest that structural breaks in time series lead to a regime switch, where each regime reflects certain behaviors that explain market dynamics over time. In practice, certain characteristics of time series can be hard to capture when time series exhibit nonlinearity, mixing, or noise. In such cases, selecting the appropriate RS model is difficult. Moreover, many of these models, especially the prevalent Markov-based switching models [27], require that the number of regimes be set prior to estimating model coefficients. This rigidity often hampers their flexibility in terms of inferring regimes from financial

*Department of Computer Science, Université de Sherbrooke, Canada, {kunpeng.xu, shengrui.wang}@usherbrooke.ca

†College of Computer and Cyber Security, Fujian Normal University, China, clfei@fjnu.edu.cn

‡Laplace Insights, Canada, jeanmarc@laplaceinsights.com

data dynamically and estimating the number of regimes.

Another prevalent limitation in RS models is that the majority of methodologies are tailored to address regime switches within an individual time series. Discovering regime switches across multiple time series is notably challenging. This complexity arises from the inherent interconnectedness within financial markets, where series do not evolve in isolation but are influenced by mutual external determinants. For instance, Fig 2(a) shows the volatility of five stocks from the S&P 500. We analyze the high-volatility region (2019-2022), which corresponds to the closing prices shown in Fig 2(b). The close prices on Carnival Corporation & plc (NASDAQ: CCL – hotels, resorts & cruise lines) decreased sharply between 2019-12 and 2020-08, while those on Tesla, Inc. (NASDAQ: TSLA – automobile manufacturers) increased significantly during the same period (as shown in Fig 2(b)-red area), which possibly indicates that people preferred private transport and refrained from travel and accommodation due to the covid pandemic. Similarly, the area outlined in blue in Fig 2(b) may be interpreted as a period of intermittent suppression of the pandemic. Exploring the linear or nonlinear interrelationships between series at different time intervals is thus crucial for regime identification and prediction.

In light of these challenges, we introduce an innovative regime-switching model tailored for multiple time series. This model emphasizes the evolving interplay between series and leverages kernel representation learning for regime identification in a nonlinear space. Our model has the following desirable properties:

(1) **Adaptiveness:** Automatically identify regimes and learn the switches, without prior knowledge about regimes.

(2) **Interpretability:** Transform heavy sets of time series into a lighter, meaningful structure via kernel representation, offering a novel perspective on regime-switching dynamics. To the best of our knowledge, there is no published work on learning kernel representation from multiple time series and accurately discovering regimes.

(3) **Effective:** Operate on multiple time series, explore the nonlinear interactions and forecast regime switching within an ecosystem. We also validate the model's efficacy in online learning scenarios. The experimental results on synthetic and real stock datasets demonstrate that our model yields superior performance in forecasting volatility.

2 Related work

The widely used regime-switching model, introduced by Hamilton [11], characterizes time series behaviors in different regimes. The model governs switches in the coefficients of an autoregression through a discrete-state Markov process. In [5, 12], the authors incorporated a switching mechanism into conditional variance models, investigated autoregressive conditional heteroskedasticity (ARCH) and generalized autoregressive conditional heteroskedasticity (GARCH) models. Chatigny et al.[6] proposed a variable-order Markov model to discover and exploit the underlying regimes in financial time series. Based on the theoretical framework of expected utility with uncertain probabilities, Wang et al. [27] proposed a Markov regime-switching model for asset pricing and stock market ambiguity measurement. Although the Markov switching model and its variants have been widely used in the analysis of economic and financial time series, they are local in the sense that one model is learned for each time series. Consequently, they cannot effectively extract information across multiple time series.

Recent literature has proposed models for addressing the question of interdependence among multiple time series. Hochstein et al. [13] proposed a multivariate smooth transition autoregression model to capture the linear interdependencies among multiple time series, in which each regime is modeled using a vector autoregressive model. It is worth noting that this type of method attempts to capture the regime shift mechanism through a single transfer matrix, which is time-dependent for series that exhibit noncontiguous regimes. Matsubara et al. [19] proposed the Regime-Cast model, which learns the various patterns that may exist in a co-evolving environment at a given window and reports the pattern(s) most likely to be observed at a subsequent time. While the approach can forecast subsequent patterns, it does not capture possible dependencies between patterns. In the updated work [20], the authors proposed the deterministic OrbitMap model for capturing the time-dependent transitions between exhibited regimes. However, in their setting, they work with priorly labeled regimes (known in advance). Numerous recent studies focus on time series analysis based on deep neural networks [23, 30, 17]; however, the primary focus of these endeavors remains on time series modeling rather than the intricate task of regime identification and understanding their persistence in financial contexts.

In light of these observations, our work aims to pioneer a fresh perspective on regime evaluation in financial time series. Eschewing traditional methods that rely on latent variable dynamics, we delve into the inherent patterns of the time series. Our proposed regime-switching framework, with its nonlinear mapping, is adept at capturing the ever-changing financial market regimes, offering insights into their intricacies and forecasting potential regime transitions.

3 Preliminaries

Notation. A time series is a set of points ordered by a time index as follows: $S_i = \{(t_l, e_l^i)\}_{l=1}^m$, where t_l denotes regular time-stamps, m signifies the series length, and e_l^i represents the series value at the specific time t_l. The index i refers to the *i-th* time series within a collection of N univariate time series $\mathbf{S} = \{S_i\}_{i=1}^N$. We denote matrices by boldface capital letters, *e.g.*, $\mathbf{M}$. $\mathbf{M}^{\mathrm{T}}$, $\mathbf{M}^{-1}$, $\mathrm{Tr}(\mathbf{M})$ indicate the transpose, inverse and trace of matrix $\mathbf{M}$, respectively. $\mathrm{diag}(\mathbf{M})$ refers to a vector with its *i-th* element being the *i-th* diagonal element of $\mathbf{M}$.

3.1 Self-representation learning in financial data Financial time series often exhibit patterns that recur over time. The concept of self-representation seeks to approximate each data point as a linear combination of other data points [18, 29]. This is expressed as $\mathbf{S} = \mathbf{S}\mathbf{Z}$ or $S_i = \sum_j S_j Z_{ij}$, where $\mathbf{Z}$ is the self-representation coefficient matrix. In the context of financial markets, if two data points S_i and S_j exhibit similar market behaviors or regimes, the coefficient Z_{ij} should be large. The learning objective function for this representation is given by:

$$\min_{\mathbf{Z}} \frac{1}{2}||\mathbf{S} - \mathbf{S}\mathbf{Z}||^2, \ s.t. \ \mathbf{Z} = \mathbf{Z}^{\mathrm{T}} \geq 0, \mathrm{diag}(\mathbf{Z}) = 0 \tag{3.1}$$

The constraint ensures the avoidance of trivial identity solutions. For financial data that is ordered based on specific market regimes, the ideal linear representation $\mathbf{Z}$ should represent each data point as a combination of points from the same market behavior or pattern. The structure of $\mathbf{Z}$ in this context can be represented as:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}^{(1)} & 0 & \cdots & 0 \\ 0 & \mathbf{Z}^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Z}^{(k)} \end{bmatrix} \tag{3.2}$$

This representation reveals the underlying structure of the time series data $\mathbf{S}$, with each block $\mathbf{Z}^{(i)}$ in the diagonal representing a specific market behavior or regime. The number of blocks k corresponds to the number of distinct market regimes.

3.2 Kernel trick for financial data modeling Linear models, while effective in many scenarios, may not always capture the intricate nonlinear relationships inherent in financial data. Kernelization techniques offer a solution by mapping data points to higher-dimensional spaces using kernel functions, allowing for linear pattern analysis in these spaces. Instead of explicitly computing the coordinates in the high-dimensional feature space, a common practice in kernel methods consists of using the "kernel trick" [14, 28]. Suppose that the nonlinear feature mapping $\Phi(\mathbf{S})$: $\mathcal{R}^d \to \mathcal{H}$ maps the data points $\mathbf{S}$ from the input space $\mathcal{R}$ to a reproducing kernel Hilbert space (RKHS) $\mathcal{H}$. It is not necessary to know the explicit representation of transformation Φ; we only need to obtain a kernel Gram matrix $\boldsymbol{\mathcal{K}} = \Phi(\mathbf{S})^\top \Phi(\mathbf{S})$. The Gaussian kernel, a popular choice, results in an infinite-dimensional feature space, making it particularly suitable for modeling complex financial data patterns [3].

4 Methodology

Financial markets are complex systems that exhibit intricate patterns over time. In this section, we propose a novel method tailored to capture the nonlinear interactions and time-evolving regimes commonly observed in financial time series data.

For the sake of clarity, consider the set of multiple co-evolving time series depicted by Fig 3(a). As highlighted in the Introduction, financial markets are characterized by nonlinear correlations and evolving regimes. The proposed method aims to capture these intricacies. We start by introducing our kernel representation learning approach, which seeks to identify homogeneous patterns through self-representation learning with a block diagonal regularizer in kernel/Hilbert space. With a given sliding window of size w, we segment time series into contiguous subseries of length w, *i.e.*, $\mathbf{S} = \bigcup_{p=1}^b \mathbf{S}_p$. The number of distinct regimes, k, is determined by counting the unique regimes across all window-stamps, $W_1, ..., W_b$. By using these identified regimes, we can then estimate regime switch probabilities. In the financial context, we define a regime as the profile pattern of a group of similar subseries observed within a specific time window. This profile pattern is essentially a subseries whose vector representation corresponds to the centroid of similar subseries, representing a specific market behavior or trend.

4.1 Modeling regime behavior To capture the intricacies of financial market regimes, we propose a kernel representation learning approach. This approach clusters subseries obtained via a sliding window technique, allowing us to identify and analyze market behaviors over different time intervals. To introduce the mathematical foundation of our method, we begin with the

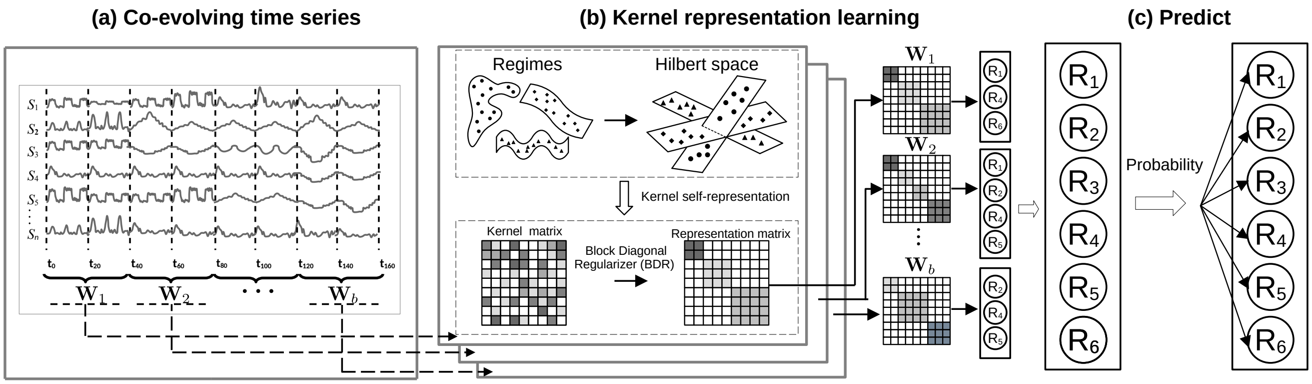


Figure 3. Framework of our method. Given multiple co-evolving time series composed of various regimes in windows, our method learns kernel representations of these regimes to predict the subsequent regimes.

simplest case, where we treat the whole series as a single window.

Given a set of time series $\mathbf{S} = (S_1, \ldots, S_N) \in \mathcal{R}^{T\times N}$ as described in Eq (3.1), its self-representation $\mathbf{Z}$ would make inner product $\mathbf{SZ}$ come close to $\mathbf{S}$ in a linear approach. However, due to the nonlinear nature of financial data, the objective (3.1) may not efficiently handle the nonlinear relations between series. To address this, we employ the "kernel tricks" to map the time series into a high-dimensional RKHS, where linear pattern analysis can be conducted. By integrating this kernel mapping, we present a new kernel representation learning strategy, as shown in Fig 3(b), with the following objective function:

$$\min_{\mathbf{Z}} ||\Phi(\mathbf{S}) - \Phi(\mathbf{S})\mathbf{Z}||^2, \ s.t.\ \mathbf{Z} = \mathbf{Z}^{\mathrm{T}} \geq 0, \mathrm{diag}(\mathbf{Z}) = 0 \tag{4.3}$$

where the mapping function $\Phi(\cdot)$ needs not be explicitly identified and is commonly replaced by a kernel $\mathcal{K}$ subject to $\mathcal{K} = \Phi(\cdot)^\top\Phi(\cdot)$.

Ideally, we hope to achieve the matrix $\mathbf{Z}$ having k block diagonals under some proper permutations if time series $\mathbf{S}$ contains k regimes. To this end, we add a regularization term to $\mathbf{Z}$ and transform Eq (4.3) to:

$$\begin{gathered}\min_{\mathbf{Z}} ||\Phi(\mathbf{S}) - \Phi(\mathbf{S})\mathbf{Z}||^2 + \gamma \sum_{i=N-k+1}^{N} \lambda_i(\mathbf{L_Z}),\\ s.t.\ \mathbf{Z} = \mathbf{Z}^{\mathrm{T}} \geq 0, \mathrm{diag}(\mathbf{Z}) = 0\end{gathered} \tag{4.4}$$

where $\gamma > 0$ defines the trade-off between the loss function and regularization terms, and $\lambda_i(\mathbf{L_Z})$ contains the eigenvalues of Laplacian matrix $\mathbf{L_Z}$ corresponding to $\mathbf{Z}$ in decreasing order. Here, the regularization term is equal to 0 if and only if $\mathbf{Z}$ is k-block diagonal (see **Theorem 1** for details). Based on the learned high-quality matrix $\mathbf{Z}$ (containing the block diagonal structure), we can easily group the time series into k regimes using traditional spectral clustering technology [22]. Note that Eq (4.4) is a nonconvex optimization problem, for which we propose a specialized method for solving the nonconvex kernel self-representation optimization (see Appendix A).

Theorem 4.1. *$\min \sum_{i=N-k+1}^{N} \lambda_i(\mathbf{L_Z})$ is equivalent to $\mathbf{Z}$ is k-block diagonal.*

Proof. Due to the fact that $\mathbf{Z} = \mathbf{Z}^{\mathrm{T}} \geq 0$, the corresponding Laplacian matrix $\mathbf{L_Z}$ is positive semidefinite, *i.e.*, $\mathbf{L_Z} \succeq 0$, and thus $\lambda_i(\mathbf{L_Z}) \geq 0$ for all i. The optimal solution of $\min \sum_{i=N-k+1}^{N} \lambda_i(\mathbf{L_Z})$ is that all elements of $\lambda_i(\mathbf{L_Z})$ are equal to 0, which means that the k smallest eigenvalues are 0. Combined with the Laplacian matrix property, the multiplicity k of the eigenvalue 0 of the corresponding Laplacian matrix $\mathbf{L_Z}$ equals the number of connected components (blocks) in $\mathbf{Z}$, and thus the soundness of **Theorem 1** has been proved. ∎

4.2 Estimating the number of financial regimes

In financial market analysis, accurately identifying the number of regimes, such as bull and bear markets or periods of varying volatility, is crucial. These regimes offer insights for informed decision-making, risk management, and strategic investment. While estimating the number of regimes in financial time series data can be challenging, our kernel representation learning approach offers a promising solution. Leveraging the block-diagonal structure of the self-representation matrix produced by our method, we can effectively estimate the number of regimes. According to the Laplacian matrix property [26], a strictly block-diagonal matrix $\mathbf{Z}$ allows us to determine the number of regimes k by first calculating the Laplacian matrix of $\mathbf{Z}$ ($\mathbf{L_Z}$) and then counting the number of zero eigenvalues of $\mathbf{L_Z}$. Recognizing that real-world financial datasets may contain noise or anomalies, which is often the case in practice, we propose an eigengap thresholding approach to estimating the number of regimes. This approach estimates the number of regimes $\hat{k}$ as:

$$\hat{k} = \arg\min_i \{i | g(\sigma_i) \leq \tau\}_{i=1}^{N-1} \tag{4.5}$$

where $0 < \tau < 1$ is a parameter and $g(\cdot)$ is an exponential eigengap operator defined as:

$$g(\sigma_i) = e^{\lambda_{i+1}} - e^{\lambda_i} \tag{4.6}$$

Here, $\lambda_{i\,i=1}^{N}$ are the eigenvalues of $\mathbf{L_Z}$ in increasing order. The eigengap, or the difference between the i^{th} and $(i+1)^{th}$ eigenvalues, plays a crucial role. According to matrix perturbation theory [24], a larger eigengap indicates a more stable subspace composed of the selected k eigenvectors. Thus, the number of financial regimes can be determined by identifying the first extreme value of the eigengap[1].

Algorithm 1 provides a comprehensive summary of the steps for nonlinear regime identification in a set of co-evolving financial time series.

4.3 Forecasting financial regime switches Given a dataset, $\mathbf{S}$, representing co-evolving financial time series, our primary objective is to model and forecast regime behaviors. This is crucial as financial markets often transition between different states, such as bullish and bearish trends. The cornerstone of our forecasting approach is the estimation of the regime-switching probability matrix.

For a set of b sliding windows, denoted as $\{W_1, \ldots, W_b\}$, we derive b kernel representations, each corresponding to a window. Suppose that, across all windows, we've identified k distinct financial regimes, *i.e.*, $\mathbf{R}_{r\in[1,k]} = \{\mathbf{R}_1, \cdots, \mathbf{R}_k\}$. It is important to know that the regimes identified within a specific window might differ when analyzed in the context of other windows (see Fig 3(b)). This reveals the variety of regimes in time series and the demand for a dynamic representation.

For two consecutive windows, W_p and W_{p+1}, we can compute the likelihood of a financial series S_i transitioning from behavior $\mathbf{R}_r$ in W_p to $\mathbf{R}_m$ in W_{p+1}, as follows:

$$P(\mathbf{R}_r \to \mathbf{R}_m | W_p \to W_{p+1}, S_i) = \frac{\sum_\star \Theta_{p,p+1}^{r,\star} \Lambda_{p,p+1}^{\star,m}}{\sum_{\star_1} \sum_{\star_2} \Theta_{p,p+1}^{\star_1,\star_2} \Lambda_{p,p+1}^{\star_1,\star_2}} \tag{4.7}$$

where $\star_1, \star_2 \in \{1, \cdots, k\}$ and

$$\begin{aligned} \Theta_{p,p+1}^{r,m} &= \frac{\eta\left(\mathbf{R}_r \to \mathbf{R}_m | \mathcal{T}r\left(S_i|W_p\right)\right)}{|\mathcal{T}r(S_i|W_p)|}, \\ \Lambda_{p,p+1}^{r,m} &= \sum_{l=1}^{p-1} \frac{\min\{\eta(\mathbf{R}_r, W_l), \eta(\mathbf{R}_m, W_{l+1})\}}{\max\{\eta(\mathbf{R}_r, W_l), \eta(\mathbf{R}_m, W_{l+1})\}} \end{aligned} \tag{4.8}$$

Here, $\eta\left(\mathbf{R}_r \to \mathbf{R}_m | \mathcal{T}r\left(S_i|W_p\right)\right)$ represents the frequency of the sequence $\{\mathbf{R}_r, \mathbf{R}_m\}$ appears in the trajectory $\mathcal{T}r\left(S_i|W_p\right)$, where $\mathcal{T}r\left(S_i|W_p\right)$ is a sequence of group to describe the various regime behavior displayed by a series S_i over time. $\eta(\mathbf{R}_r, W_l)$ (resp. $\eta(\mathbf{R}_m, W_{l+1})$) is the number of series presenting behavior $\mathbf{R}_r$ (resp. $\mathbf{R}_m$) at window W_l (resp. W_{l+1}).

Note that the component $\Theta_{p,p+1}^{r,m}$ gauges the risk of observing a sudden shift to behavior $\mathbf{R}_m$ given the prior behavior was $\mathbf{R}r$. Meanwhile, $\Lambda_{p,p+1}^{r,m}$ assesses the overall likelihood of transitioning between the two regimes within the entire dataset $\mathbf{S}$. Thus, our approach considers both the immediate risk associated with a single financial series and the broader context of the market.

To forecast the series value of S_i in W_{p+1}, given its behavior in W_p is $\mathbf{R}_r$, and the most probable regime transition leads to $\mathbf{R}m$, we can estimate the value based on historical series and the predicted state. The forecasted values of S_i for the window $Wp+1$ are given by:

$$Forecast\_S_i = \sum_{l=1}^{p} \Delta(\mathbf{R}_m | S_i, W_l) \cdot \tau^{p-l+1} \cdot \{S_i | W_l\} \tag{4.9}$$

where the indicator function $\Delta(\mathbf{R}_m|S_i, W_l)$ indicates whether S_i belongs to $\mathbf{R}_m$ under window W_l, and $\{S_i|W_l\}$ is the subseries value of S_i within window W_l. The weight value $\tau^{p-l+1} \in (0,1)$ modulates the contribution of past data to the forecast, emphasizing the importance of recent financial data.

5 Experiments

5.1 Data To evaluate our model, we used two real-world stock datasets. The first of these, `Stock1`, comprises 503 stocks collected from the S&P 500, composed of daily OHCLV (open, high, close, low, volume) data from 2012-01-04 to 2022-06-22[2]. The second, `Stock2`, is composed of intra-day market hours OHCLV data from 2017-05-16 to 2017-12-06, for 467 stocks from the S&P500[3]. Our primary objective was

[1] We initialize the number of regimes for each window as 3 to obtain the initial representation $\mathbf{Z}$, and then estimate k.

[2] `https://ca.finance.yahoo.com/`

[3] `https://www.kaggle.com/datasets`

Algorithm 1 Nonlinear regime identification

Data: A set of multiple time series $\mathbf{S} = \{S_i\}_{i=1}^N$, kernel $\boldsymbol{\mathcal{K}}$, window size w;
Result: Number of regimes k, set of regimes $\mathbf{R}$, set of representaions $\mathcal{Z} = \{\mathbf{Z}_p\}_{p=1}^b$;

```
begin
    - w ← Window size
    /* Scanning and Learning:                */
    begin
        - Z ← ∅ - At each window W_p get the set of
          subseries S_p - Obtain Z_p, using Eq (4.4);
        - Z ← Z ∪ {Z_p}
    /* Discovering regimes:                  */
    - Based on the Z, R ← distinct regimes, as mentioned
      in section 4.1 and k ← |R|
```

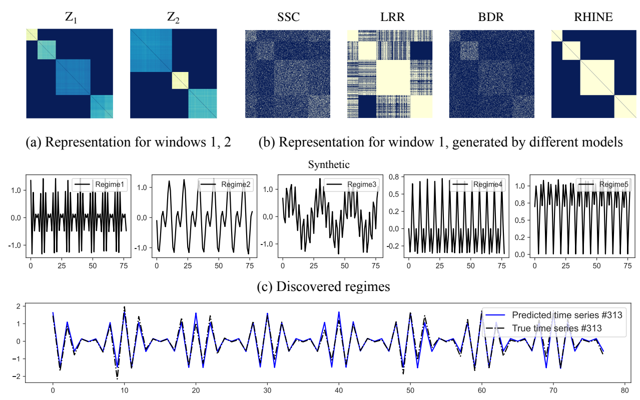
to predict the implied volatility of each stock[4]. Given that true volatility remains elusive, we approximated it using an estimator grounded in realized volatility. We employed the conventional volatility estimator, defined as: $\mathcal{V}_t = \sqrt{\sum_{t=1}^n (r_t)^2}$, where $r_t = \ln(c_t/c_{t-1})$ and c_t represents the closing price at time t. For `Stock1`, we utilized daily data to gauge monthly volatility, while for `Stock2`, we used 1-hour intra-day data to determine daily volatility.

In addition to real data, we crafted a `Synthetic` dataset comprising 500 simulated volatility time series, each generated by a combination of 5 nonlinear functions (see Appendix B for details). The synthetic dataset allows the controllability of the structures/numbers of regimes and the availability of ground truth. For data generation, we randomly selected one of the 5 functions ten times. Each selection yielded 78 sequential values, which were treated as a regime.

5.2 Experimental setup and evaluation For the kernel representation learning phase, we opted for the Gaussian kernel, defined as $\boldsymbol{\mathcal{K}}(S_i, S_j) = exp(-||S_i - S_j||^2/d_{max}^2)$, with d_{max} as the maximum series distance. Parameters γ in Eq (4.4) and β in Eq (7.10) (found in Appendix A) were selected from [0.1,0.4,0.8,1,4,10] and [5,10,20,40,60,100], respectively. The optimal values were determined to be $\gamma = 0.8$ and $\beta = 60$.

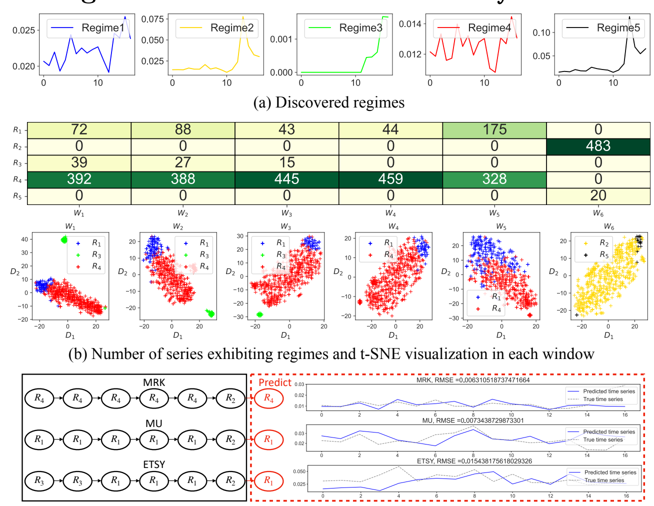
To verify the effectiveness of the proposed RHINE model, we evaluated the volatility forecasting performance of RHINE against seven different models.

[4]We chose to validate our regime-switching model through forecasting as it provides a quantifiable and objective measure to assess the model's capability to understand and adapt to market changes, rather than through investment decisions or predicting market trends, which could be influenced by subjective interpretations and external market conditions and fall outside the purview of this study.



(a) Representation for windows 1, 2 (b) Representation for window 1, generated by different models
(c) Discovered regimes
(d) True and predicted time series (*e.g.*, series#313)

Figure 4. Visualized results on the `Synthetic`



(a) Discovered regimes
(b) Number of series exhibiting regimes and t-SNE visualization in each window
(c) Examples of regime prediction and regime-based series forecasts for subsequent window

Figure 5. Visualized results on the `Stock1`

Among them, four are forecasting models (ARIMA [4], KNNR [7], INFORMER [30], and a state-of-the-art ensemble model N-BEATS [23]), the other three are RS models (MSGARCH [1], SD-Markov [2] and OrBitMap [20]). For the existing methods, we use the codes released by the authors, and the details of the parameter settings can be found in Appendix C. We evaluated the models' performance via the Root Mean Square Error (RMSE).

5.3 Regime identification and prediction To evaluate the capability of RHINE to identify the regimes, we used the `Synthetic` and `Stock1` datasets. For real case, we do not have a ground truth for validating the obtained regimes. Instead, we resort to an indirect method, *i.e.*, comparing the volatility values based on the forecast regime with the true volatility values computed from the original time series. In the learning phase, a fixed window traverses all series, generating subseries under varying windows. Subsequently, we derive a kernel representation for the subseries in each window. The quality of our kernel representation thus depends highly on how well the time series is split. Our method for automatically estimating the optimal size of the sliding windows to obtain suitable regimes is detailed in Appendix D. The method allows obtaining window sizes of 78 and 17 for `Synthetic` and `Stock1`, respectively (see Fig 10 in Appendix D).

Result on `Synthetic` With the determined optimal window size, RHINE learns the kernel representations for all subseries within each window. Fig 4(a) illustrates the self-representation for 500 subseries of the synthetic time series for the initial two windows, *N.b.*: each element of this 500×500-dimensional representation matrix signifies the correlation between two subseries within a window. It is evident that subseries with higher correlation are more congruent, as indicated by a brighter point. The varying representations in the two windows reveal that the correlation between subseries is time-dependent, – *i.e.*, different regimes are concealed in these two windows. Fig 4(b) depicts the representation (binarized version) for window 1, as generated by RHINE and three other state-of-the-art self-representation learning models – *i.e.*, SSC [10], LRR [16], and BDR [18]. RHINE yields a block diagonal matrix with dense within-regime scatter and sparse between-regime separation, effectively revealing the true underlying regime structure (note that indistinct regimes can lead to bad representations). By plotting the profile pattern of each block as shown in Fig 4(c) we can see that the 5 regimes precisely correspond to the 5 functions used to generate the data. Fig 4(d) presents the forecasts for the last window of a synthetic time series, demonstrating the proximity of the forecasts to the actual values. This underscores that tracking the series interaction over time can help with time series forecasting for subsequent times.

Result on `Stock1` Fig 5(a) showcases the diverse regimes manifested by co-evolving series in `Stock1`, implying multiple patterns underlying the monthly volatility. With the identified regimes, we can discern series that exhibit each of these regimes at different window-stamps, aiding in the comprehension of regime switching. For illustrative purposes, Fig 5(b) displays the heatmap of each regime at varying windows and their corresponding 2D visualizations using t-Distributed Stochastic Neighbor Embedding (t-SNE) [25]. The heatmap reveals that series indeed harbor different regimes at distinct window-stamps, with the color intensity indicating the prevalence of the regime. For instance, the gradual lightening of the color in the third row of the heatmap from left to right indicates the diminishing observation of the corresponding regime in most series over time. By analyzing the identified regimes in Fig 5(a) and correlating them with the heatmap in Fig 5(b), it's evident that $\mathbf{R}_1$ and $\mathbf{R}_4$ are the most prevalent regimes, representing standard market volatility and normal trading activities. The exclusive display of $\mathbf{R}_2$ and $\mathbf{R}_5$ in the 6-th window signifies substantial market fluctuations, – *i.e.*, the COVID-19 pandemic has severely impacted the financial markets. In Fig 5(c), we pick three stocks (NASDAQ: MRK, MU, and ETSY) from `Stock1` and present how the regime switches and how the RHINE predicts the series values based on the predicted regime at the next window. It can be seen that the predicted values fit well with the original time series. Tracking the regime over time clearly helped in predicting series values at subsequent times, although the predicted values at some time intervals may deviate from the true time series.

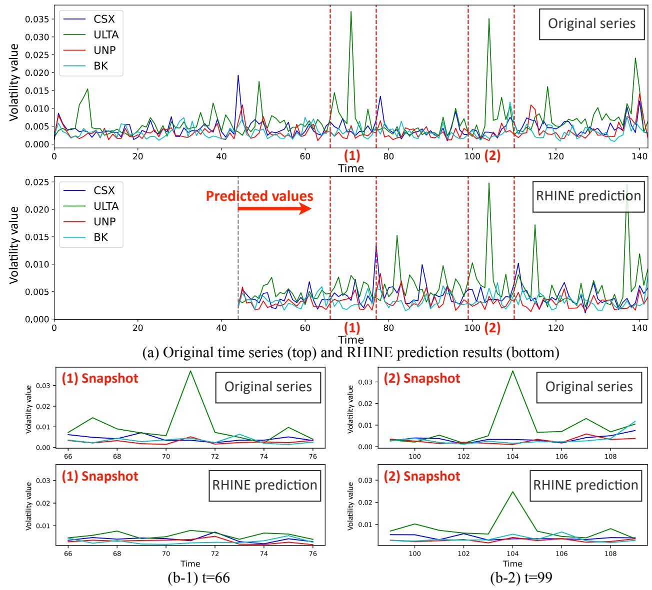


Figure 6. Online forecasting results on the `Stock2` time series

5.4 Online learning on `Stock2`

To further illustrate the predictive capabilities of RHINE, we employed it for one of the most challenging tasks in time series analysis – *i.e.*, online forecasting, leveraging the discovered regimes. For this task, we need to forecast upcoming unknown future events, at any given moment, while discarding redundant information. This approach is inherently aligned with online learning paradigms, where the model continually learns and adapts to new data points, making it highly pertinent in the dynamic landscape of financial markets. We conducted tests on the `Stock2` dataset, segmenting the time series with an 11-day sliding window. Given the constrained intra-day data available (7 months) and the strategy employed for volatility forecasting, extending the sliding window

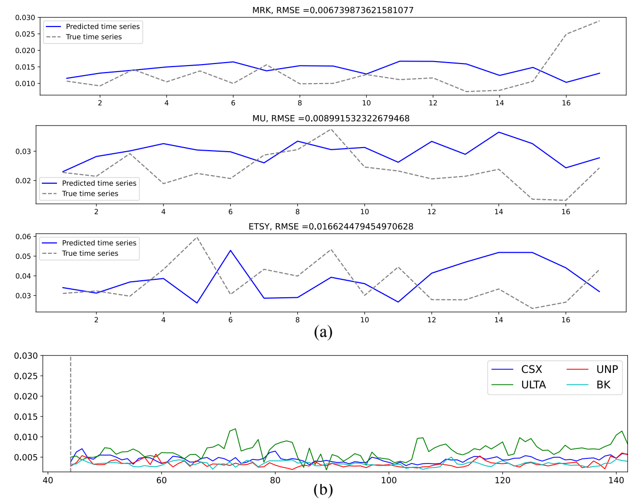


Figure 7. Results of N-BEATS model. (a) forecasting on three series of `Stock1`. (b) online forecasting on four series of `Stock2`.

would compromise the available data for assessment. At any time point t, the observable data encompasses a period quadruple the window size preceding time t, *e.g.*, at time point t=110, RHINE is trained with $\mathbf{S}\,[66:109]$ and forecasts $\mathbf{S}\,[110:121]$.

Fig 6 delineates online forecasting examples on four stocks (NASDAQ: CSX, ULTA, UNP, and BK) and some snapshots obtained at several different timestamps. The original data at the top of Fig 6(a) elucidates the daily volatility fluctuations for these four stocks. The lower part of Fig 6(a) unveils the outcomes for online forecasting, showcasing how RHINE anticipates series behavior over time. For stock ULTA, depicted by the green line, all the compared models, including RHINE, encounter challenges in predicting the abnormal behavior of the first high volatility (Fig 6(b-1)) due to the absence of antecedent knowledge. However, post encountering this anomalous behavior, RHINE accurately anticipates the timing of the second anomalous high-volatility behavior (Fig 6(b-2)) and the ensuing volatility behavior, attributing to RHINE's ability to model regime switching by leveraging the interrelations among multiple time series. Specifically, for a single-series regime-switching model, it is impossible to predict the second anomalous behavior accurately if there is no periodic pattern in it; whereas for RHINE, since our regime-switching is based on correlations between series, when other series start to show some anomalous volatility behavior (albeit small), RHINE is also able to predict the next volatility of multiple time series in a holistic way.

5.5 Benchmark comparison

Table 1 delineates the comparative forecasting efficacy of the various models. It is evident that RHINE consistently surpasses its counterparts, registering minimal forecasting error across all datasets. ARIMA is capable of capturing seasonality within series; however, it encounters challenges in apprehending complex, nonlinear dynamic interactions amidst financial time series when various seasonalities are disjointed. In most cases, N-BEATS (the SOTA forecasting model) emerges as the second-best owing to the advantages of ensemble models. Fig 7 displays the actual forecasting outcomes of N-BEATS. Compared to our forecasted result, shown in Fig 5(c) and Fig 6, N-BEATS is unsuitable for capturing complex regime switches within multiple time series; it is a single time series forecasting model and is not able to forecast abrupt changes of regimes. OrbitMap yields acceptable results, as it is also a regime-aware method, but underperforms RHINE from the need for predefined regimes and difficulty with multiple time series. Fig 8 compares RHINE with other methods in terms of computation time in linear-log scales indicating the great time efficiency of RHINE only outperformed by OrbitMap.

Table 1. Models' forecasting performance, in terms of RMSE, for the three datasets (`Stock2` reports the mean value of online forecast sliding)

Models		Datasets		
		`Synthetic`	`Stock1`	`Stock2`
Forecasting models	ARIMA	1.761	2.635×10^{-2}	2.918×10^{-2}
	KNNR	1.954	2.348×10^{-2}	2.761×10^{-2}
	INFORMER	0.672	1.214×10^{-2}	1.006×10^{-2}
	N-BEATS	0.319	1.035×10^{-2}	6.074×10^{-3}
RS models	MSGARCH	1.264	2.366×10^{-2}	2.129×10^{-2}
	SD-Markov	0.936	2.146×10^{-2}	1.669×10^{-2}
	OrbitMap	0.635	1.003×10^{-2}	7.469×10^{-3}
	RHINE	**0.315**	$\mathbf{8.780\times10^{-3}}$	$\mathbf{3.032\times10^{-3}}$

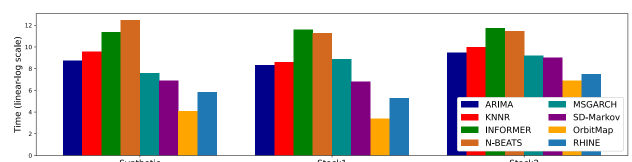


Figure 8. Computation time on the three data sets

6 Conclusion

This paper introduces RHINE, a novel regime-switching model, adept at uncovering and modeling nonlinear interactions within an ecosystem of multiple time series, particularly in financial markets. RHINE enables us to model regimes explicitly by extracting inherent patterns through a newly devised kernel representation of time series. The resulting time-varying kernel representation matrices are instrumental in identifying regime switches. Significantly, RHINE autonomously reveals diverse concealed regimes, eliminating the necessity for prior knowledge about the investigated series. Our experimental results demonstrated that RHINE surpasses existing models in forecasting accuracy, providing a robust solution in the financial domain for volatility forecasting and real-time analysis of stock market dynamics.

References

[1] D. Ardia, K. Bluteau, K. Boudt, L. Catania, and D.-A. Trottier, *Markov-switching garch models in r: The msgarch package*, Journal of Statistical Software, 91 (2019).

[2] M. Bazzi, F. Blasques, S. J. Koopman, and A. Lucas, *Time-varying transition probabilities for markov regime switching models*, Journal of Time Series Analysis, 38 (2017), pp. 458–478.

[3] G. Benton, W. Maddox, and A. G. Wilson, *Volatility based kernels and moving average means for accurate forecasting with gaussian processes*, in International Conference on Machine Learning, PMLR, 2022, pp. 1798–1816.

[4] G. Box, *Box and jenkins: time series analysis, forecasting and control*, in A Very British Affair, Springer, 2013, pp. 161–215.

[5] J. Cai, *A markov model of switching-regime arch*, Journal of Business & Economic Statistics, 12 (1994), pp. 309–316.

[6] P. Chatigny, R. Chen, J.-M. Patenaude, and S. Wang, *A variable-order regime switching model to identify significant patterns in financial markets*, in 2018 IEEE International Conference on Data Mining (ICDM), IEEE, 2018, pp. 887–892.

[7] R. Chen and I. Paschalidis, *Selecting optimal decisions via distributionally robust nearest-neighbor regression*, Advances in Neural Information Processing Systems, 32 (2019).

[8] R. Chen, M. Sun, K. Xu, J.-M. Patenaude, and S. Wang, *Clustering-based cross-sectional regime identification for financial market forecasting*, in International Conference on Database and Expert Systems Applications, Springer, 2022, pp. 3–16.

[9] J. Dattorro, *Convex optimization & Euclidean distance geometry*, Lulu. com, 2010.

[10] E. Elhamifar and R. Vidal, *Sparse subspace clustering: Algorithm, theory, and applications*, IEEE TPAMI, 35 (2013), pp. 2765–2781.

[11] J. D. Hamilton, *A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle*, Econometrica, 57 (1989), pp. 357–384.

[12] J. D. Hamilton and R. Susmel, *Autoregressive conditional heteroskedasticity and changes in regime*, Journal of econometrics, 64 (1994), pp. 307–333.

[13] A. Hochstein, H.-I. Ahn, Y. T. Leung, and M. Denesuk, *Switching vector autoregressive models with higher-order regime dynamics application to prognostics and health management*, in PHM, IEEE, 2014, pp. 1–10.

[14] Z. Kang, C. Peng, and Q. Cheng, *Twin learning for similarity and clustering: A unified kernel approach*, in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31-1, 2017.

[15] Z. Lin, R. Liu, and Z. Su, *Linearized alternating direction method with adaptive penalty for low-rank representation*, Advances in neural information processing systems, 24 (2011).

[16] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, *Robust recovery of subspace structures by low-rank representation*, IEEE TPAMI, 35 (2012), pp. 171–184.

[17] Y. Liu, H. Wu, J. Wang, and M. Long, *Non-stationary transformers: Exploring the stationarity in time series forecasting*, Advances in Neural Information Processing Systems, 35 (2022), pp. 9881–9893.

[18] C. Lu, J. Feng, Z. Lin, T. Mei, and S. Yan, *Subspace clustering by block diagonal representation*, IEEE TPAMI, 41 (2018), pp. 487–501.

[19] Y. Matsubara and Y. Sakurai, *Regime shifts in streams: Real-time forecasting of co-evolving time sequences*, in ACM SIGKDD, 2016, pp. 1045–1054.

[20] ——, *Dynamic modeling and forecasting of time-evolving data streams*, in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 458–468.

[21] T. D. Mount, Y. Ning, and X. Cai, *Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters*, Energy Economics, 28 (2006), pp. 62–80.

[22] A. Ng, M. Jordan, and Y. Weiss, *On spectral clustering: Analysis and an algorithm*, Advances in neural information processing systems, 14 (2001).

[23] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, *N-beats: Neural basis expansion analysis for interpretable time series forecasting*, arXiv preprint arXiv:1905.10437, (2019).

[24] G. W. Stewart, *Matrix perturbation theory*, (1990).

[25] L. Van der Maaten and G. Hinton, *Visualizing data using t-sne.*, Journal of machine learning research, 9 (2008).

[26] U. Von Luxburg, *A tutorial on spectral clustering*, Statistics and computing, 17 (2007), pp. 395–416.

[27] J. Wang, M. Zhou, X. Guo, L. Qi, and X. Wang, *A markov regime switching model for asset pricing and ambiguity measurement of stock market*, Neurocomputing, 435 (2021), pp. 283–294.

[28] K. Xu, L. Chen, and S. Wang, *Data-driven kernel subspace clustering with local manifold preservation*, in 2022 IEEE International Conference on Data Mining Workshops (ICDMW), IEEE, 2022, pp. 876–884.

[29] ——, *A multi-view kernel clustering framework for categorical sequences*, Expert Systems with Applications, 197 (2022), p. 116637.

[30] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, *Informer: Beyond efficient transformer for long sequence time-series forecasting*, in Proceedings of the AAAI conference on artificial intelligence, vol. 35, 2021, pp. 11106–11115.