Dynamic Cross-sectional Regime Identification for Financial Market Prediction

Anonymous

Abstract—**We investigate issues related to dynamic cross-sectional regime identification for financial market prediction. A financial market can be viewed as an ecosystem regulated by particular regimes that may switch at different time points due to structural breaks driven by market forces in the financial ecosystem. Most existing regime-based prediction approaches assume that the training data were sufficiently representative of possible regimes occurring in the market, which prevent them from identifying new regimes in the process of prediction since the regimes only switches between a fixed number of already-identified regimes with a static transition probability matrix. Such an assumption is unrealistic and prevents these approaches from being effective for real-world applications since financial markets are time varying and may fall into a new regime at any future time. Moreover, most of such approaches are focused on single time series. These shortcoming prompted us to devise a dynamic cross-sectional regime identification model for time series prediction. The new model is defined on a multi-time-series system, with time-varying transition probabilities that can identify new cross-sectional regimes dynamically from the time-evolving financial market. Experimental results on real-world financial datasets illustrate the promising performance and suitability of our model.**

Index Terms—**Dynamic Regime identification, Cross-sectional Regime identification, Financial Market Prediction**

I. Introduction

A critical need in financial market prediction is the capacity to foresee market behavior in periods of high or low stress on financial time-series, and to establish links between market dynamics and the general state of the market. This is an important topic that has been investigated by many researchers and practitioners due to its potential financial gain [1], [2]. However, studying the market behaviors underlying financial data is not an easy task because certain market behaviors, characterized by dynamic patterns, known as regimes, are not directly observable, such as skewness and fat tails, downside risk properties, and time varying correlations [3], especially for those characterized by the fat tail distribution, volatility clustering, asymmetry and mean reversion [4]. In such cases, modeling the behavior of financial data and making prediction still remains an opening research problem.

Regime switch analysis is extensively advocated to capture some aspects of complex behaviors and continues to gain popularity in many fields for tasks, such as distinguishing between woodlands and grasslands in an ecological system [5], walking and wiping activities in motion sensors [6] and weekday and weekend consumption in energy systems [7]. A regime can be defined by a specific group of linear or/and non-linear patterns that share identical dynamics or other characteristics at a specific time, where a regime may switch to another within that ecosystem. Regime switch is interpreted here as occasional, discrete shifts in the parameters governing the dynamic behavior caused by external drivers and/or internal feedbacks [8]. For example, Fig. 1 shows a time series in an ecosystem, regulated by three regimes: colored red (R_1), blue (R_2) and yellow (R_3), respectively, which consist of three groups of distinct patterns and contiguously repeat over time in the ecosystem.

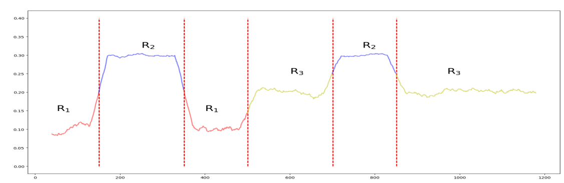


Fig. 1. An example of an ecosystem dominated by three regimes colored red (R_1), blue (R_2) and yellow (R_3), respectively. Vertical dashed lines delimit the regimes over time.

Recent advances in regime switch models [9], [10], [11] make them a promising approach in financial market studies to explore dynamic patterns or characteristics of financial time series for market prediction. Most existing regime models are designed to predict the likelihood of a structural break resulting in a regime switch that is driven by a combination of driving variables, pressures from within or outside the market [8], [12], [13]. However, a common drawback of these models is that they are based on the assumption that the training data were sufficiently representative of possible regimes occurring in the market with a static transition probability matrix, which prevents these methods from identifying new regimes in the process of prediction due to the fact that the regimes only switches between a fixed number of already-identified regimes. For example, if the data from before the 2008 financial crisis are used to train a model, the model will likely predict the 2008 financial crisis regime as one of the regimes identified previously and can not identify the crisis regime of 2008. Thus, it would be difficult to achieve an acceptable result due to the fact that the market may switch to new regimes at any time, and the transition probabilities also vary over time and depend on some underlying economic fundamentals as they are affected by current market conditions.

On the other hand, some models such as the Markov switching models have also been proposed to allow the transition probabilities to vary over time by using observable covariates, including strictly exogenous explanatory variables and lagged values of the dependent variables [14]. However, the performance of these approaches depends on application-specific prior knowledge such as which variables or which functions are used to describe the dynamics of the transition probabilities [9]. Moreover, these models still suffer from the issue of switching between a fixed number of already-identified regimes due to the fact that they are not capable of dynamically identifying new regimes occurring in the time-evolving market. Such limitations prevent these models from being effective for real-world applications since financial markets are time varying and may fall into a new regime at any future time. For example, the 7^{th} time interval (the last one) in Fig. 2 is not included in the training data, existing regime models will predict it as the closest already-identified regime R_1. However, the actual regime is a new one (R_4), different from all the previously identified regimes (R_1, R_2, R_3), which results in poor performance of the prediction model.

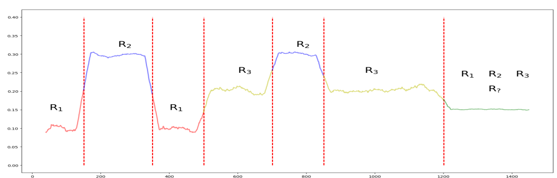


Fig. 2. An example of an ecosystem dominated by three regimes prior to time step 1200. Note that a new regime occurred after time step 1200 in the time evolving market. Vertical dashed lines delimit the regimes over time.

Inspired by the challenges above, we propose a model for dynamic cross-sectional regime identification in financial data for market prediction. Our model allows identifying the number of cross-sectional regimes dynamically according to the regime identification process in the time-evolving financial market, to bypass the problem of regime switching among a fixed set of regimes with a static transition probability matrix. Moreover, the regime transition probability of our dynamic regime identification model is time-varying over the dynamic market, which provides insights for financial market behavior analysis increases the interpretability of the model's explanations. The major contributions of this work can be summarized as follows:

- We propose a model for dynamic cross-sectional regime identification in financial data for market prediction, which allows identifying the number of regimes dynamically during the regime identification process in the time-evolving financial market, to eliminate the drawback of switching into already-identified regimes with a static transition probability matrix.
- We propose a time-varying transition probability over the dynamic market, which provides insights for financial behavior analysis in the markets and better interpretability for model predictions.
- We propose a new data reconstruction methods by combining three different volatility intervals to capture dynamic patterns, such as monthly, quarterly and annual dynamic patterns, for regime identification in financial markets.
- We validate our model by implementing it on multi-time series financial data. We illustrate the suitability of the proposed method by comparing its performance with that of the baseline algorithms. Comprehensive experimental results demonstrate the promising performance of our model.

The remainder of this paper is organized as follows. In section II, we discuss related work on the market prediction. Section III presents the proposed dynamic cross-sectional regime identification based prediction model in details. Section IV provides empirical results on financial data and compares the results with other baselines. Finally, conclusions are given in section V.

II. Related Work

Recent literature reports a number of approaches applied to financial market prediction, typically using statistical models, such as the AR, ARIMA [1], TBATS [2], VARMA [15] and GARCH families [16]. However, these approaches are based on the assumption that the residuals are uncorrelated and normally distributed. If either of these assumptions does not hold, then the prediction may be incorrect. Moreover, they are all based on linear equations, and are thus incapable of modeling financial data governed by non-linear dynamics. To model the non-linear dynamics of financial market behaviors, deep neural network based methods have been proposed for market prediction [17], [18]. while these approaches can achieve relatively good prediction performance, it comes at very high computational cost because of the overwhelming number of parameters. Moreover, from the standpoint of model interpretability, they are unable to provide an effective account for the regime-switching mechanism.

To render the model interpretable, Hamilton [19] proposed a regime switch model for financial market prediction in which the regime switches are governed by unobservable state variables, following a fixed-order Markov chain process. Deng et al. [20] proposed a regime switch model to capture the price spikes using a continuous-time framework. Huisman et al. [21] proposed a three-regime model on electricity price data that is able to predict regime switches using constant Markov transition probabilities. Ethier and Mount [22] proposed a regime switch model for the behavior of electricity spot prices in various deregulated markets. Klaassen et al. [23] proposed a two-regime Markov regime-switch GARCH model on volatility to make multi-period ahead volatility forecasting a convenient recursive procedure. Marcucci et al. proposed a regime-switching GARCH model to forecast stock market volatility in financial markets [24]. However, all of these approaches assume that the regime transition probabilities are constant over time, which is impractical for real application since the probabilities vary over time and are affected by changing market conditions.

As mentioned earlier, Markov techniques have been used to allow the transition probabilities to vary over time by using observable covariates that include strictly exogenous explanatory variables and lagged values of the dependent variable. Marco et al. [14] proposed a Markov switch model allowing transition probabilities to vary over time as specific transformations of lagged dependent observations. Luca et al. [25] proposed a Markov regime switch model using time-varying transition probabilities to adjust US fiscal policy for asset prices. Lee et al. [26] proposed a regime switch model with time-varying transition probabilities for optimal investment. Those approaches can be useful and effective, on the condition that users specify what variables or functions are to be used to describe the dynamics of transition probabilities [9]. Moreover, they assume that all the regimes occurring in the test data are the same as the regimes identified in the training data. In other words, they are incapable of dealing with the situation depicted in Fig. 2.

In this paper, we propose a dynamic regime identification model with time-varying transition probability. Not only our model identifies novel regimes via the regime identification process, but also it is capable of predicting switches to these new regimes by the time-evolving financial market. It can provide meaningful insights for financial market behavior analysis and powerful model interpretability as a result of the regimes' time-varying transition probabilities determined over the dynamic financial market.

III. The Proposed Model

In this section, we give a detailed description of our dynamic model for cross-sectional regime identification and prediction, from data pre-processing to the final market prediction, which includes volatility calculation, data reconstruction, model description and estimation.

A. Overview of the Proposed Model

This subsection gives an overview of the proposed cross-sectional regime prediction model through the scenario depicted in Fig 3. While our cross-sectional model is designed for multiple time-series data, we present it on data containing a single time series for a concise explanation from the model description standpoint, to provide insights and a better understanding of the mechanisms underlying the model. We start from identifying the first regime R_1 from the first slide window and make a prediction, and then implement the regime identification and prediction process iteratively. In a word, we need to identify the regime of the time series segment X_w, in order to make a prediction of e^v, here $w = 126$ a variable corresponding to the duration of the current regime while $v = 21$, usually prefixed, corresponds to the number of future times stamps for the prediction. Thus, assume that K regimes $(R_k, 1 \leq k \leq K)$ have been identified before time $t - w$; followed K estimations of $e_i^w (1 \leq i \leq K)$. By comparing the slide window X_w with the K estimations e_i^w, if the smallest error is greater than $\eta||X_w||$ where $\eta = 0.5$, a new regime will be identified based on the current X_w, and added to the regime database R_{db} and obtained the transition probability p_{K+1}; otherwise we identified the regime of current window X_w, and add X_W to the identified regime and update the regime parameters, and then update the transition probability P. Finally, we recognized the current regime R_{p_m} by maximal regime transition probability p_m and make a prediction $e^v = e^v_{R_m}$. The overview of the dynamic cross-sectional regime prediction model is shown in Fig 3.

B. Volatility Calculation

The point of interest for market prediction is to predict the implied daily volatility of the stocks or indexes as the true volatility is time evolving. The literature reports many approaches for volatility calculation, such as equally or weighted approaches [27], [28]. However, the main idea is to estimate the volatility by the variance of log returns over an interval of trading days, where the log return, presented in [29], is defined as the log of the ratio of closing prices at adjacent points in time as follows:

$$r_t = \ln \frac{p_t}{p_{t-1}} \tag{1}$$

where p_t is the closing price at time t, and based on the daily return, the classical volatility estimator presented in [28] is estimated by the variance of log returns over each interval with m trading days as follows:

$$\sigma_{t+1} = \sqrt{\frac{1}{m} \sum_{s=t-m}^{t} (r_s - \mu)^2} \tag{2}$$

where r_s is the logarithm daily return of the s^{th} trading day. m specifies the number of trading days at the beginning of day t and μ is the mean of m trading days.

However, the volatility in Eq. (2) assigns equal weights to the amounts of historical data, which means that all the data in the m trading days are of equal importance in reflecting the dynamics of the data. In practice, however, due to the non-stationary nature of financial data, the most recent historical data are more relevant to reflect the time-evolving dynamics and give the most information for future prediction. Thus, in order to capture the dynamics patterns of time series for a better prediction, the exponentially weighted moving average approach was proposed to capture the time relevant movements and dynamic patterns in the volatility of a time series by assigning much more weight to the most recent data using an exponential function [27]. The exponentially weighted volatility can be estimated as follows:

$$\sigma_{t+1} = \sqrt{(1-\lambda) \sum_{s=t-m}^{t} \lambda^{t-s-1} (r_s - \mu)^2} \tag{3}$$

where λ is the weight decay factor, by which the past data decay as they become more distant, and $\sum_{i=1}^{m} \lambda^i = 1$. Note that, the sum of the weights will be approaching but not to equal 1 in practice since only limited historical data will be used during the implementation A large decay factor implies a value-at-risk measure that is derived almost exclusively from

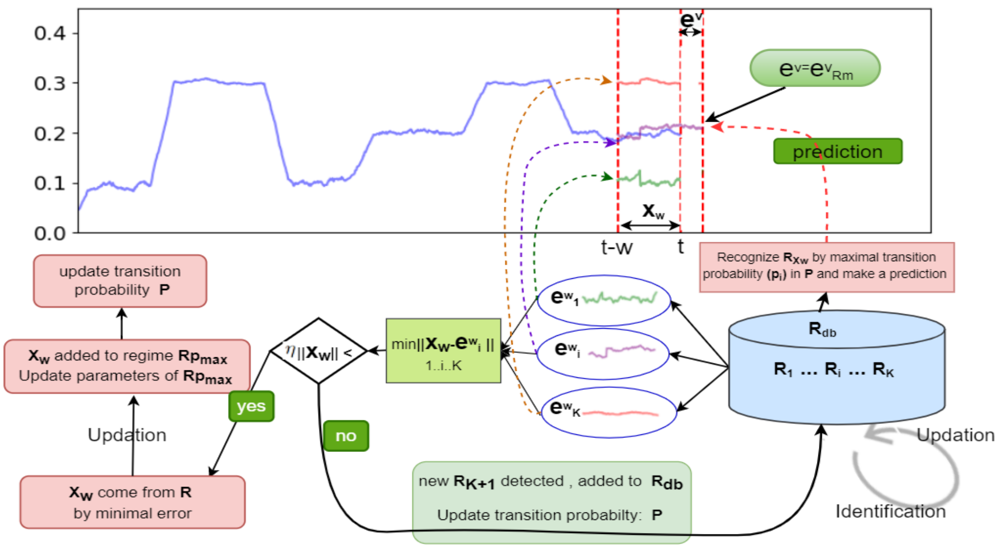


Fig. 3. An overview of the dynamic cross-sectional regime based prediction model.

very recent data. We therefore introduced it into this paper, and set $\lambda = 0.8$ in the implementation, while setting $\mu = 0$ as Figlewski presented in [30].

C. Data Reconstruction

The basic purpose of data reconstruction is to make the data "talk a lot", which can help us explore the dynamic patterns hidden in the data, such as financial market dynamics or behaviors. To achieve a better prediction, we hope to make use of all possible time-evolving dynamic patterns in the financial markets, both the short- and long-term, e.g.,monthly, quarterly and annual patterns. However, to explore the diversity patterns, we use a multi-layer approach different time volatility intervals, to capture different layer dynamic patterns in the volatile financial markets. The longer the interval, the less fluctuation in volatility, due to the fact that the volatility will be smoothed by longer duration. For example, we can use 3 different volatility intervals ({21, 63, 126}) to capture the monthly, quarterly and annual dynamic patterns in the volatile financial markets, respectively, and then reconstruct these into a new one that may be characterized by much more dynamics patterns in the market than each one individually. The reconstruction is defined as follows:

$$\sigma_t = \sum_{i=1}^{m} \omega_i \times \sigma_t^i \tag{4}$$

where σ_t^i is the i^{th} volatility interval derived from the exponentially weighted volatility by Eq. (3). In this paper, we use 3 ($m = 3$) different interval ({21, 63, 126}) to capture the monthly, quarterly and half annual dynamical patterns in volatile financial markets. Thus, this approach can capture both short-term and long-term market behaviors in the time evolving financial system. Here, we assign following weights ($\omega \in \{0.25, 0.5, 0.25\}$) to the three different intervals to balance the regime identification and prediction. And detailed explanation is provided in the Dataset Description sub-section.

D. Regime Analysis and Market Prediction

This subsection contains regime analysis and market prediction. The regime analysis includes a detailed description of the regime model and regime identification.

1) Model Description: Inspired by the model proposed in [5], [6], [31], we propose a dynamic cross-sectional regime identification model for volatility prediction in a financial market regulated by different regimes. For a better model description, we assume that there are K identified regimes in the dynamic financial market consisting of multiple time series. Thus, the market can be defined as follows:

$$\frac{ds_k(t)}{dt} = a_k + \mathcal{G}_k g(s_k(t)) + \mathcal{F}_k f(s_k(t)) \quad (1 \leq k \leq K) \tag{5}$$

$$\frac{d\Delta_k(t)}{dt} \propto p_k(t) \tag{6}$$

$$e(t) = \underset{p_1,\cdots,p_K}{\arg\max}\, \epsilon_k + \mathcal{E}_k s_k(t) \tag{7}$$

where $ds(t)/dt$ denotes the derivative with respect to time t, $s(t)$ denotes the potential activities in market, and $e(t)$ the estimated values. $g(\cdot)$ is a linear function, while $f(\cdot)$ is non-linear. Here, a, $\mathcal{G}$, $\mathcal{F}$ describe the potential activities $s(t)$, capturing linear, exponential, and non-linear dynamic patterns of the financial ecosystem respectively. $\Delta_k(t)$ is the transition contributor of switching to k^{th} regime at time t, and $p_k(t)$ is the probability of switching to the k^{th} regime at time t and P is a regime switch matrix, while ϵ, $\mathcal{E}$ are the observation projection that are used to estimate the $e(t)$ based on the identified regime. Therefore, a regime R can be described by $\Theta = \{a, \mathcal{G}, \mathcal{F}, \epsilon, \mathcal{E}, p\}$, while the dynamics financial markets with K regimes ecosystem can be expressed as $\Re = \{R_1, \ldots, R_K, P\}$.

2) Regime Identification: In order to identify the regimes in a financial market, we need to estimate the optimal parameters which describe the dynamic patterns of each regime. Here, we will give details of the parameter estimation process on one regime. We divide the regime parameter R into two parts: a linear part $\Theta_L = \{a, \mathcal{G}, \epsilon, \mathcal{E}\}$, and a non-linear part $\Theta_N = \{\mathcal{F}\}$, as presented in [6]. These parts describes the linear and nonlinear dynamic patterns of the financial market, respectively, and can be optimized separately by the expectation-maximization (EM) algorithm.

To obtain the optimal parameters of regime (R), we initialize the nonlinear parameter Θ_N: $\mathcal{F} = 0$ for linear parameter Θ_L estimation. We can then estimate e on the initialized Θ_N for the E step of the EM algorithm. The process of estimating the linear parameter Θ_L can be described as follows:

$$\begin{gathered} e = \psi(s_0^{'}, \Theta_L^{'}, \Theta_N) \\ \{s_0, \Theta_L\} = \underset{s_0^{'}, \Theta_L^{'}}{\arg\min} \parallel X_R - e \parallel \end{gathered} \tag{8}$$

where X_R denotes all the data belonging to regime R, $\psi(\cdot)$ is a function that uses to estimate e. Note that, in contrary to the estimation process presented in [6], our regime identification process is based on the whole historical data with respect regimes, since history data are of significance for future prediction in financial markets. After obtaining the linear parameter Θ_L, we can get e on the Θ_L for the nonlinear parameter Θ_N estimation, which is called the M step of the EM algorithm, and the estimation process for the nonlinear parameter Θ_N can be described as follows:

$$\begin{gathered} e = \psi(s_0^{'}, \Theta_L, \Theta_N^{'}) \\ \{s_0, \Theta_N\} = \underset{s_0^{'}, \Theta_N^{'}}{\arg\min} \parallel X_R - e \parallel \end{gathered} \tag{9}$$

The next step is the transition probability estimation P, which can be defined as follows:

$$\begin{gathered} e_k^w = \psi(s_0^k, \Theta_L^k, \Theta_N^k) \qquad \Delta_k(t) = \ln ||X_w - e_k^w|| \\ p_k(t) = \frac{\exp(\frac{d\Delta_k(t)}{dt})}{\sum_{k=1}^{K} \exp(\frac{d\Delta_k(t)}{dt})} \quad P = P \cup \{p_k(t)_{k=1}^K\} \end{gathered} \tag{10}$$

Therefore, regime R can be described by $\Theta = \{a, \mathcal{G}, \mathcal{F}, \epsilon, \mathcal{E}, p\}$, while the dynamics financial ecosystem with K regimes can be expressed as $\Re = \{R_1, \ldots, R_K, P\}$.

3) Market Prediction: Once all the financial ecosystem parameters $\Re = \{R_1, \ldots, R_K, P\}$ have been estimated, the next step is volatility prediction on this system. Actually, the process of obtaining the regime parameters is the regime identification process. We can identify the regime of current time point with the maximal transition probability, we then can make a prediction based on the identified regime, the prediction is defined as follows:

$$e^v = \underset{p_1, \ldots, p_K}{\arg\max} \, \psi(s_0^k, \Theta_L^k, \Theta_N^k) \tag{11}$$

The framework of the prediction process is shown in detail in **Algorithm 1**.

IV. Experiments

In this section, we provide a description of the test data sets and then introduce the evaluation methods used to validate the performance of our proposed model. Finally, we compare the empirical results with other baselines.

A. Datasets

To test the performance of our model, we choose 9 extensively used financial markets Sectors: the one consists of nine financial sector SPDR Fund: XLB, XLE, XLF, XLI, XLK, XLP, XLU, XLV, XLY; the other consists of 12 stocks selected from top 500 companies: AAPL, IBM, BAC, MSFT, MS, WMT, INTC, C, CVX, JPM, HIG, WFC. Both of two datasets consist of daily frequencies (for only business days), comprising over 22 years' data (a total of 5626 business days) including Dot-com Bubble (2001-2002), the financial crisis (2008-2009), and the covid-19 (2019-2020) period from Jan 3^{th}, 1999 to May 31^{th}, 2021. These data are available on the google Finance website [1]. We used the data reconstruction method for data processing from the log return to the reconstructed data as shown in Fig. 4. Comparing with the four volatilities, it can be easily seen that the reconstructed volatility (blue line) delays the changes in monthly volatility (red line) and speeds up the changes in annual volatility (green line). Moreover, it is much smoother than the quarterly volatility (gray line). To some extent, this reconstruction method provides a better description of the complex dynamic patterns in financial markets, such as earlier structural breaks, which can be used for market regime identification and switch analysis.

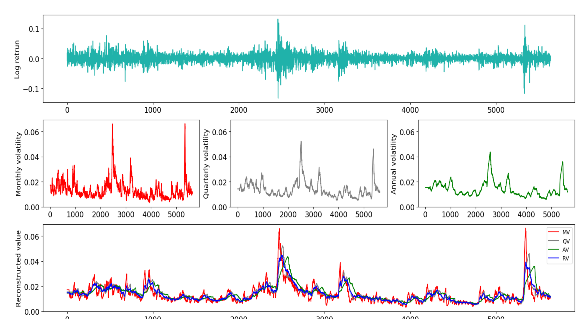


Fig. 4. An example of reconstructed volatility from log return of XLB sector

B. Evaluation Methods

There are several different measurements for evaluating volatility prediction performances. To evaluate the prediction ability, we chose two of the most commonly used loss functions the mean square error (RMSE),which is based on a

[1] https://ca.finance.yahoo.com/

Algorithm 1: Framework of proposed model

```
Input: Financial data: X, Slide window Length: w,
Prediction window: v
Threshold η
Output: Predicted volatility: E
begin
    /* Data processing */
      • Log-return calculation;
      • Volatility estimation ;
      • Data transformation;
    /* Initialization */
      • tc current time point;
      • ℜ = [], regimes;
      • E = [], predicted time series;
      • TP = [], transition probability;
    repeat
        /* Get window data X_w from X */
        X_w = X[tc-w:tc];
        if len(ℜ) ==0 then
            /* Identify regime on X_w */
            ℜ.append({R});
            /* Estimated value on R */
            e^v = ψ(R);
            /* predicted value on R */
            E.append(e^v);
            /* p_1(tc) = 1.0 */
            TP.append(p_R(tc));
        err = [];
        for R in ℜ do
            e^w_R = ψ(R);
            err.append(sqrt(||e^w_R − X_w||));
            /* obtain p_R(tc) */
            TP.append(p_R(tc));
        if min(err) > η||X_w|| then
            /* identify new R_new on X_w */
            ℜ.append(R_new);
            /* obtain p_{R_new}(tc) on err */
            TP.append(p_{R_new}(tc));
        else
            /* identified regime of X_w */
            X_w ∈ R_m;
            /* Update TP on err */
            TP.append(p_{R_m}(tc));
            Insert X_w to R_m;
            Update R_m parameters;
        R_pm = arg min_{p_1,··· ,p_i,···} || ℜ ||
        e^v_{R_m} = ψ(R_pm);
        E.append(e^v);
    until move for next window;
```

quadratic loss function, and the mean absolute error (MAE), which is less sensitive to severe mispredictions than the RMSE [32]. We also used the mean absolute percentage error (MAPE) introduced by Bollerslev and Ghysels [33]. The loss functions are as follows:

$$MAE = \frac{1}{T}\sum_{t=1}^{T} |\sigma_i - \hat{\sigma_i}| \tag{12}$$

$$MSE = \frac{1}{T}\sum_{t=1}^{T} (\sigma_i - \hat{\sigma_i})^2 \tag{13}$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T} (\sigma_i - \hat{\sigma_i})^2} \tag{14}$$

$$MAPE = \frac{1}{T}\sum_{t=1}^{T} |\frac{\sigma_i - \hat{\sigma_i}}{\sigma_i}| \tag{15}$$

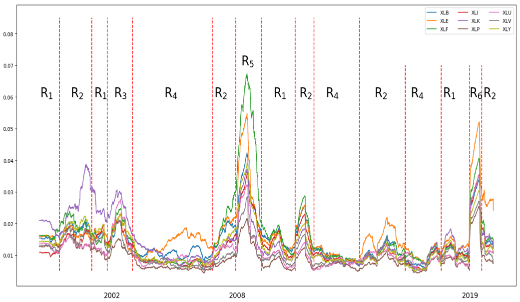
where T is the length of the rolling window, σ_i, $\hat{\sigma_i}$ are the actual and predicted volatility at time t respectively. The smaller of the values, the closer are the predicted time series values to the actual values, and the better performance of the predicting model.

C. Experimental Results and Discussion

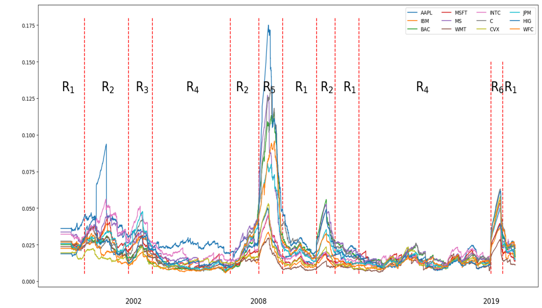
In this section, we present the results of our model, and evaluate its performance against some comparable methods on the nine sectors that comprising the financial market also called as the financial ecosystem. Since our model is designed to perform predictions based on the cross-sectional regime identification for financial market. We tested our model with 21 days ahead of prediction. The results are presented in the following sub-sections.

1) Cross-sectional Regime Identification Analysis: Our model is designed for cross-sectional regime identification based market prediction, so we get start from regimes identified on financial ecosystem and perform a regime analysis. First, we present the regimes identified by our model for the ecosystem consisting of two datasets as shown in Fig. 5. It is clear that 6 different regimes, i.e. R_1, R_2, R_3, R_4, R_5 and R_6 delimited by dashed lines, were identified with respect to two ecosystems as shown in Fig. 5(a) and Fig. 5(b), which means that both of these two financial ecosystems are dominated by 6 different regimes. However, as shown in two sub-figures in Fig. 5, it can be observed that these regimes are distributed differently in time, and the switching times are not synchronous in the two ecosystems, in some cases, a switch to the same regime occurs at the same time. Comparing the two sub figures, we can observe some concordance between regime intervals. However, there exist also distinct intervals, for instance, corresponding to the first regime R_2 interval in Fig. 5(b), there exist two regimes R_2 and R_1 in Fig. 5(a); and corresponding to the second regime R_4 interval in Fig. 5(b) there exist 3 different regimes (R_1, R_2 and R_4) in Fig. 5(a)

On the other hand, by looking at the common regimes of two ecosystems in Fig. 5, we observe also some identical regimes



(a) Regimes identified in sector ecosystem.



(b) Regimes identified in stock ecosystem.

Fig. 5. Regimes identified in sector and stock ecosystem. In each of case, 6 regimes (R_1, R_2, R_3, R_4, R_5 and R_6) were identified. Regimes are delimited by vertical dashed lines

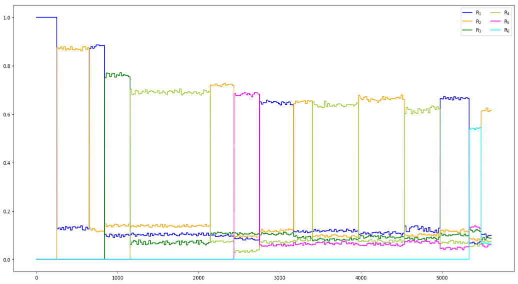
with the same transitions at the same time points. Moreover, we notice that regimes R_3, R_5 and R_6 occur only once in the whole market, while the others are repeated at least twice; and if they are combined with big events in the financial market, the timestamps of the three regimes (R_3, R_5 and R_6) roughly coincide with the Dot-com Bubble (2001-2002), the financial crisis (2008-2009), and the covid-19 (2019-2020) period, respectively, which confirms that our model is capable of identifying cross-sectional regimes in the financial market.

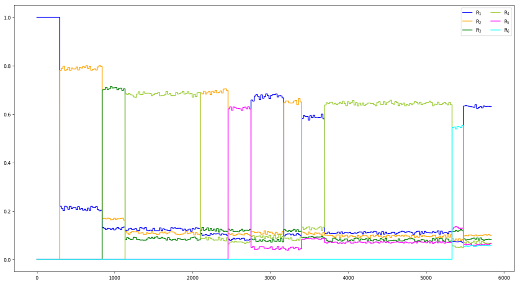
2) Time Varying Transition Probability Analysis: The key problem in regime switch analysis is when and where to switch. The transition probability can define the likelihood of remaining in the current regime or being forced to switch to another. Thus, the transition probability is of great significance to the regime identification and regime switch analysis. Since our regime identification model is a dynamic model whose regime database is expanded dynamically once a new regime arrives in the market. As noted previously, this is completely different from existing regime models, in which the only switches possible are among the already-identified regimes, with a static transition probability, and new regimes cannot be identified as they arrive.

As presented above, we identified 6 regimes in the sector and stock ecosystem: R_1, R_2, R_3, R_4, R_5 and R_6. The time-varying transition probabilities of the two ecosystems with 6 regimes are shown in Fig. 6(a) and Fig. 6(b). For a better explanation of the regime switch analysis, we show them individually as in Fig. 6(c) and Fig. 6(d) respectively. We will exhibit the regime transition process based on the time-varying transition probabilities shown in Fig. 6. We take the transition probabilities of sector ecosystem Fig. 6(a) as our example for a further analysis of the regime switching process. There is only one regime R_1 in the market at the beginning, and the transition probability of R_1 is set to 1 (blue line). We assume that the transition probabilities to non occurred regimes are set to 0 (the other five are set to 0), while regime R_2 was identified, the transition probabilities of regime R_2 is greater than R_1 as shown in Fig.6(c). Thus, the transition probabilities to the other 4 regimes are set to 0.

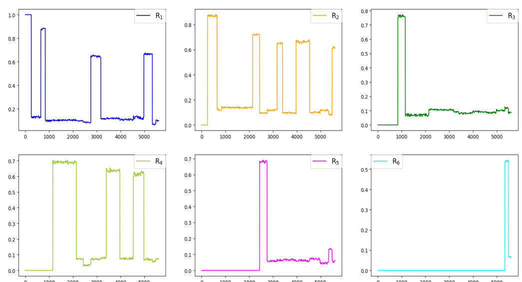
Actually, this is an unrealistic assumption due to the fact that the regime is not observable and the number of regimes is also unknown, and the regime transition probability is incalculable until the corresponding regime turns up in the market. We are adopting this setting here for a better explanation from the market standpoint. In fact, the transition probability is an indicator forcing the regime switching or not. Looking at the transition probability of regime R_6 (cyan line) switched from regime R_1 (blue line) in Fig. 6(a), the transition probability of regime R_5 (magenta line) is largest, followed by R_3 (green line), R_1 (blue line) and R_4 (yellowgreen line). This means the structures or dynamic patterns of regime R_6 is dissimilar from regime R_4 where there should be a regime switch. It is much similar to that of regime R_5, but it is still different from regime R_5. Thus a new regime was identified and the system force to switch to a new regime R_6, and this can be easily confirmed in Fig. 5 by comparing with the regime structures of R_1, R_2, R_3 and R_4. In a word, this kind of time-varying transition probability not only provides a way to analyze the dynamic regime-switching behaviors hidden in financial markets, but also helps us understand the regime-switching mechanism from model explanation standpoint, which is completely different from the existing regime switching model with a static regime transition probability.

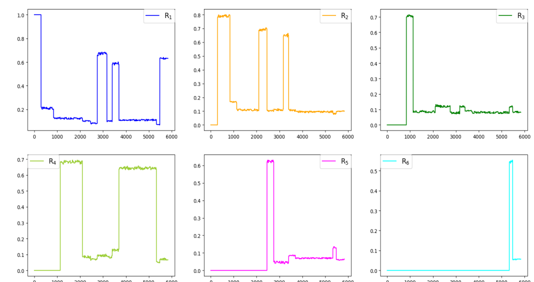
3) Market Prediction Analysis: To show the prediction quality of our model, we use a slide window with 126 business days and make a prediction with 21 days ahead of time on two real-world datasets. However, the results of the baselines were omitted in the figures, as their high error values make it hard to display our prediction values and errors legibly in the same figure. However, we compared our results with the baselines in terms of average error using the performance metrics MAE, MSE, RMSE and MAPE later. The predicted value and ground truth of two datasets are shown Fig. 7(a) and Fig. 7(b), respectively. It is hard to find the difference between the predicted value and ground truth, we thus present them on individually for a better view of our prediction shown in Fig. 7(c) and Fig. 7(d). It is clear that our predicted values





(a) Time-varying transition probability among 6 regimes in sector ecosystem. (b) Time-varying transition probability among 6 regimes in stock ecosystem.





(c) Time-varying transition probability of each regime in sector ecosystem. (d) Time-varying transition probability of each regime in stock ecosystem.

Fig. 6. Time-varying transition probability switching among 6 regimes in sector and stock ecosystem are shown in Fig. 6(a) and Fig. 6(b). For a better explanation for regime switching analysis based on the transition probabilities, we shown them individually as in Fig. 6(c) and Fig. 6(d) respectively.

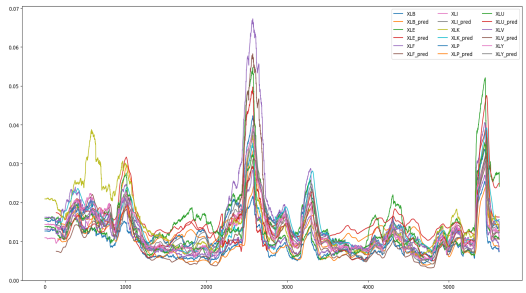
TABLE I
COMPARISON OF MAE, MSE, RMSE AND MAPE RESULTS ON THE TWO DATASETS. THE BEST RESULTS ARE IN BOLD.

	Sectors				Stocks			
	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
ARIMA	32.416	1.168	11.485	1621.538	42.318	1.567	14.321	1989.053
VARMA	26.715	0.876	9.873	1574.399	38.926	1.216	12.494	1882.981
ICA-GARCH	23.823	0.694	8.534	1469.351	36.591	0.984	11.367	1755.365
MTGNN	19.681	0.467	6.923	1301.905	34.762	0.813	8.136	1601.632
DeepGlo	17.638	0.359	4.329	1228.484	30.569	0.572	7.834	1532.514
RegimeCast	20.513	0.511	7.397	1391.952	32.214	0.784	9.593	1682.169
Our model	**13.887**	**0.129**	**2.946**	**1064.282**	**26.715**	**0.311**	**5.829**	**1474.399**

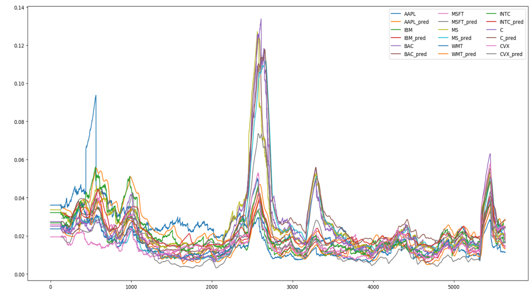
(red lines) are well matched by the ground truth (blue lines) on every sub-pictures, which means our model exhibit good performance on the prediction. The four time-evolving errors with respect to MAE, MSE, RMSE, MAPE on two ecosystem during the prediction process are shown in Fig. 7(e) and Fig. 7(f).

4) Performance Analysis: In this sub-section, we will show the prediction quality of our model, and estimate its prediction accuracy relative to the some baselines on the same datasets. We used six benchmark models to compare the multi-step forecasting performance in the two ecosystem: ARIMA [1] and VARMA [15] are the classical statistical models for analyzing and forecasting time series data. We determined their optimal parameter by using AIC. ICA-GARCH [34] uses independent component analysis (ICA) for transforming the multivariate time series into statistically independent time series for financial market prediction. DeepGlo [35] is a deep neural network approach to high-dimensional time series forecasting. MTGNN [36] is used to forecast multivariate time series with graph neural networks. RegimeCast [6] is a regime-based stream prediction model. The four error metrics MAE, MSE, RMSE, MAPE for the prediction process are shown in Table. I.

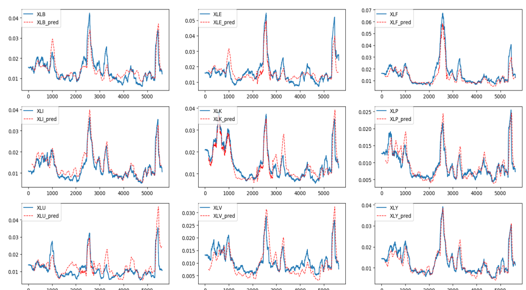
We can see that our model outperforms the methods and earns the smallest values on each performance metrics, and shows a significant improvement than the second best method DeepGLO in Table I. ARIMA and VARMA have poor performance since they are linear models and can not deal with nonlinear dynamics for financial time series data. ICA-GARCH need to transforming the multivariate time series into



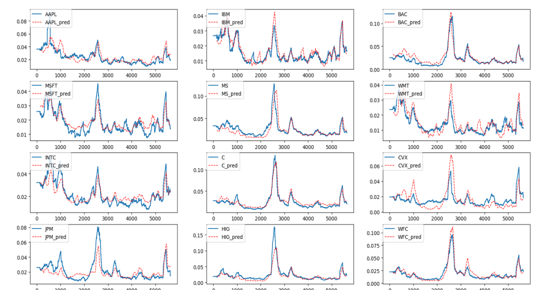
(a) Prediction vs ground truth of sector ecosystem.



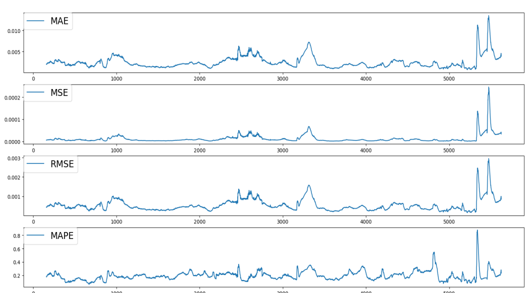
(b) Prediction vs ground truth of stock ecosystem.



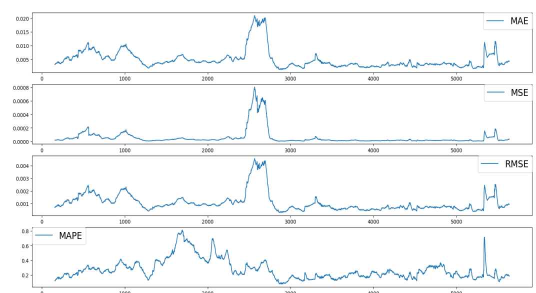
(c) Prediction vs ground truth on sector ecosystem shown individually.



(d) Prediction vs ground truth on stock ecosystem shown individually.



(e) Time-varying average error of sector ecosystem.



(f) Time-varying average error of stock ecosystem.

Fig. 7. The predicted value and ground truth with 21 days ahead are shown in Fig. 7(a) and Fig. 7(b), while Fig. 7(c) and Fig. 7(d) show individually for a better view of the prediction. The time-evolving errors with respect to MAE, MSE, RMSE and MAPE during the prediction process are shown in Fig. 7(e) and Fig. 7(f).

statistically independent time series, which of course inferior these neural network based on models MTGNN and DeepGLO, but still better than ARIMA and VARMA. MTGNN and DeepGLO can capture the nonlinear dynamics, but they focus on short-term prediction and show pool performance on long-term prediction, while DeepGLO still achieves the second place. However, RegimeCast only relies on the data within a slide window and ignores the importance of historical data for the prediction. In summary, our model shows good performance on both regime detection and financial time series prediction.

V. Conclusion

In this paper, we proposed a dynamic cross-sectional regime identification model for financial market prediction, which is designed to operate by identifying the cross-sectional regime patterns present in financial data with time-varying transition probability. Our dynamic model is not restricted to switch in a set of already-identified regimes, but also can switch to a new regime, thanks to its dynamic regime identification mechanism that can detect new regimes based on current dynamic changes. Moreover, its unique time varying transition probability provides a better explanation of dynamic market behaviors resulting in regime switching, which is completely different from existing regime identification models with static transition probability. Experimental results on financial data demonstrate the promising performance of our model.

Our future work will focus on the following aspects: we will investigate correlations between individual financial data based on our model, which, to some extent, can help to provide insights for monetary policymakers and practitioners seeking to analyze the time-varying dynamics and financial behaviors hidden in markets and predict upcoming structural changes of the volatility analysis for financial markets, helping them make better decisions to reduce risk and maximize profits for their investment. In short, we see significant challenges for our future work, but we are confident that our method has great potential in real applications.

Acknowledgment

No available

References

[1] G. E. Box and D. A. Pierce, "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models," *Journal of the American statistical Association*, vol. 65, no. 332, pp. 1509–1526, 1970.

[2] A. M. De Livera, R. J. Hyndman, and R. D. Snyder, "Forecasting time series with complex seasonal patterns using exponential smoothing," *Journal of the American statistical association*, vol. 106, no. 496, pp. 1513–1527, 2011.

[3] A. Ang and A. Timmermann, "Regime changes and financial markets," *Annu. Rev. Financ. Econ.*, vol. 4, no. 1, pp. 313–337, 2012.

[4] C. Alexander, *Market models: A guide to financial data analysis*. John Wiley & Sons, 2001.

[5] M. Scheffer, S. Carpenter, J. A. Foley, C. Folke, and B. Walker, "Catastrophic shifts in ecosystems," *Nature*, vol. 413, no. 6856, p. 591, 2001.

[6] Y. Matsubara and Y. Sakurai, "Regime shifts in streams: Real-time forecasting of co-evolving time sequences," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 1045–1054.

[7] F. Fallahi, "Causal relationship between energy consumption (ec) and gdp: a markov-switching (ms) causality," *Energy*, vol. 36, no. 7, pp. 4165–4170, 2011.

[8] M. Barange, G. Beaugrand, R. Harris, R. I. Perry, M. Scheffer, F. Werner *et al.*, "Regime shifts in marine ecosystems: detection, prediction and management," *Trends in Ecology & Evolution*, vol. 23, no. 7, pp. 402–409, 2008.

[9] R. T. Baillie and C. Morana, "Modelling long memory and structural breaks in conditional variances: An adaptive figarch approach," *Journal of Economic Dynamics and Control*, vol. 33, no. 8, pp. 1577–1592, 2009.

[10] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[11] T. Teräsvirta, "An introduction to univariate garch models," in *Handbook of Financial time series*. Springer, 2009, pp. 17–42.

[12] L. Charfeddine and K. B. Khediri, "Financial development and environmental quality in uae: Cointegration with structural breaks," *Renewable and Sustainable Energy Reviews*, vol. 55, pp. 1322–1335, 2016.

[13] A. Banerjee and G. Urga, "Modelling structural breaks, long memory and stock market volatility: an overview," *Journal of Econometrics*, vol. 129, no. 1-2, pp. 1–34, 2005.

[14] M. Bazzi, F. Blasques, S. J. Koopman, and A. Lucas, "Time-varying transition probabilities for markov regime switching models," *Journal of Time Series Analysis*, vol. 38, no. 3, pp. 458–478, 2017.

[15] H. Lütkepohl, "Forecasting with varma models," *Handbook of economic forecasting*, vol. 1, pp. 287–325, 2006.

[16] G. Ali *et al.*, "Egarch, gjr-garch, tgarch, avgarch, ngarch, igarch and aparch models for pathogens at marine recreational sites," *Journal of Statistical and Econometric Methods*, vol. 2, no. 3, pp. 57–73, 2013.

[17] A. Kim, Y. Yang, S. Lessmann, T. Ma, M.-C. Sung, and J. E. Johnson, "Can deep learning predict risky retail investors? a case study in financial risk behavior forecasting," *European Journal of Operational Research*, vol. 283, no. 1, pp. 217–234, 2020.

[18] L. O. Orimoloye, M.-C. Sung, T. Ma, and J. E. Johnson, "Comparing the effectiveness of deep feedforward neural networks and shallow architectures for predicting stock price indices," *Expert Systems with Applications*, vol. 139, p. 112828, 2020.

[19] J. D. Hamilton, "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica: Journal of the Econometric Society*, pp. 357–384, 1989.

[20] S. Deng, *Stochastic models of energy commodity prices and their applications: Mean-reversion with jumps and spikes*. University of California Energy Institute Berkeley, CA, 2000.

[21] R. Huisman and R. Mahieu, "Regime jumps in electricity prices," *Energy economics*, vol. 25, no. 5, pp. 425–434, 2003.

[22] R. Ethier and T. Mount, "Estimating the volatility of spot prices in restructured electricity markets and the implications for option values," PSerc Working Paper 98-31, Tech. Rep., 1998.

[23] F. Klaassen, "Improving garch volatility forecasts with regime-switching garch," in *Advances in Markov-Switching Models*. Springer, 2002, pp. 223–254.

[24] J. Marcucci, "Forecasting stock market volatility with regime-switching garch models," *Studies in Nonlinear Dynamics & Econometrics*, vol. 9, no. 4, 2005.

[25] L. Agnello, G. Dufrénot, and R. M. Sousa, "Using time-varying transition probabilities in markov switching processes to adjust us fiscal policy for asset prices," *Economic Modelling*, vol. 34, pp. 25–36, 2013.

[26] H.-C. Lee, S. Park, and J. M. Yoon, "Optimal investment with time-varying transition probabilities for regime switching," *Journal of Derivatives and Quantitative Studies: Seonmul yeon'gu*, 2021.

[27] B. J. Christensen and N. R. Prabhala, "The relation between implied and realized volatility," *Journal of financial economics*, vol. 50, no. 2, pp. 125–150, 1998.

[28] J. C. Hull, *Options futures and other derivatives*. Pearson Education India, 2003.

[29] M. B. Garman and M. J. Klass, "On the estimation of security price volatilities from historical data," *Journal of business*, pp. 67–78, 1980.

[30] S. Figlewski, "Forecasting volatility," *Financial markets, institutions & instruments*, vol. 6, no. 1, pp. 1–88, 1997.

[31] Y. Matsubara and Y. Sakurai, "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.

[32] S. Makridakis, "Accuracy measures: theoretical and practical concerns," *International Journal of Forecasting*, vol. 9, no. 4, pp. 527–529, 1993.

[33] T. Bollerslev, R. F. Engle, and D. B. Nelson, "Arch models," *Handbook of econometrics*, vol. 4, pp. 2959–3038, 1994.

[34] E. H. Wu and L. Philip, "Volatility modelling of multivariate financial time series by using ica-garch models," in *International Conference on Intelligent Data Engineering and Automated Learning*. Springer, 2005, pp. 571–579.

[35] R. Sen, H.-F. Yu, and I. Dhillon, "Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting," *arXiv preprint arXiv:1905.03806*, 2019.

[36] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, and C. Zhang, "Connecting the dots: Multivariate time series forecasting with graph neural networks," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020.