21AIE461

Higher Order Data Driven Techniques for Financial Forecasting

TEAM 7

ARUN PRAKSH - CB.EN.U4AIE19014
C.R. ASSWIN - CB.EN.U4AIE19016
AVINASH DORA - CB.EN.U4AIE19028
SAI GANESH - CB.EN.U4AIE19044
SIVA PRAKASH - CB.EN.U4AIE19060

I.Abstract:

Financial stock data forecasting is a challenging task due to the inherent stochasticity in the model. We propose a method for forecasting based on Higher-Order Dynamic Mode Decomposition with the assumption that the stock market is a dynamic system. DMD, being equation-free and with a low computation cost is compared with the existing ARIMA technique. The results obtained from DMD forecasting outperform the traditional approaches. HODMD was used to capture the information that DMD had missed out showing a better forecast compared to DMD. An iterative strategy was used to fix the window length for the DMD and HODMD approaches. The effect of intra and inter-dependencies in the stock market was also analyzed by modeling based on stocks belonging to the same sector and different sectors respectively. The data used in this study was extracted from the Nifty stock exchange and the results are compared based on the mean absolute percentage error. The results show that HODMD outperforms the existing statistical techniques for stock price forecasting.

Index Terms—Higher Order Dynamic Mode Decomposition, Dynamic Mode Decomposition, Mean Absolute Percentage Error.

II.Introduction:

Stock price prediction is important to estimate and make decisions to increase the returns. However, stock market prediction is not an easy task due to characteristics such as high volatility and unpredictability. The technical analysis of the stock market states that the current stock price is dependent on the previous day's stock price. The fundamental analysis of the stock market states that stock predictions can be made by analyzing the current financial situation and by taking other external factors into account. Most technical analysis is done using statistical techniques such as moving average, autoregressive, autoregressive moving average, and autoregressive integrated moving average. While the autoregressive method uses past data for future prediction, and moving average uses past errors to determine future predictions, ARMA uses both the error and the previous data for forecasting.

However, the limitations of such techniques lie in the changing dynamics of the model which affect the long-term predictions. Predictions using neural networks were considered to capture the pattern in the data. The limitation of neural network-based forecasting is the need for training on huge data which might not capture all the possible patterns that the market might exhibit.

Our work proposes using HODMD to understand the patterns in the data. HODMD, an extension of the dynamic mode decomposition was used to identify the spatiotemporal structures in the data and to find the best fit line for regression. The advantage of using the proposed method lies in its low complexity. We propose using HODMD for both short-term and long-term predictions and compare the performance of the proposed model against the ARIMA technique.

The rest of this article is organized as follows: Section 2 describes the literature survey. The theoretical background is discussed in Section 3. Section 4 details the methodology. The experimental results are analyzed and discussed with plots in Section 6. In Section 7, we conclude our work; along with the possible future works.

III. Related Works

The applications of DMD gained popularity across various domains with the initial beginning in the fluid dynamics community. One of the most well-known applications is data-driven modeling for complex systems [1,2]. Financial markets have inherent properties of stochasticity and follow no laws of nature, unlike fluids. Studies [3,4,5] show the advantages of using complex systems to understand the dynamics of such markets. The application of DMD for financial trading strategies was introduced by Mann et al. [6]. The authors use DMD to decompose the complex system into low-rank structures that capture the spatio-temporal information in its sampling windows. The extracted information is then used to forecast the data using the best-fit regression line. The authors use an iterative strategy to finalize the optimal combination of the sampling window and the number of future forecast days to get the highest returns. Kuttichira et al. [7] proposed forecasting using three different strategies. The first strategy is to sample companies belonging to the same sector to predict future stock prices. The second strategy is to sample companies belonging to different sectors to predict the future of stock prices. The third strategy is prediction until a threshold value. The results conclude that the financial market exhibits interdependencies across different sectors and the predictions for any sector which was modeled considering all other sectors prove beneficial compared to modeling on only one sector. In [8], the authors incorporate other external influences in addition to the application of DMD. The trading window is frozen when the success of the prediction rate is less than 50%. Higher-Order Dynamic Mode Decomposition (HODMD) was proposed to access the information that was previously not accessible by using DMD in [9]. The adaptation of HODMD from fluid dynamics to heart disease analysis is explained in [10]. This work proposes using HODMD for stock price prediction and analyzing the effects of intra and interdependencies in the Indian stock market.

IV.Theoritical Background :

4.1. Dynamic Mode Decomposition

Dynamic mode decomposition is a versatile and strong matrix decomposition technique that originated in fluid dynamics and nonlinear waves and was developed by peter Schmid. It was developed in order to comprehend, control, or simulate inherently complex, nonlinear systems without necessarily knowing the underlying governing equations that drive the system completely or partially. When we see DMD in the last few years alone, it has made immense progress in both theory and application. The DMD algorithm provides spatio-temporal decomposition of data into its corresponding dynamic modes via the snapshots of a given system and has a connection to non-linear dynamics of the system through the Koopman operator which helps in mapping time-delayed snapshots. DMD produces a least-square regression to reduce the high-dimensional system to a lower-dimensional system in the form of DMD modes and its corresponding eigenvalues. The DMD modes represent the coherent spatial structure of the data and its corresponding eigenvalues show the system's evolving behavior. There are two methods for obtaining these eigenvalues and modes. The first is Arnoldi-like, which is useful for theoretical analysis due to its connection with Krylov's methods. The second is a singular value decomposition (SVD) based approach that is more robust to noise in the data and to numerical errors. Stock market data is considered complex dynamical systems and the stock prices are interactive and can be considered as a measurement of such complex systems. By extracting the dynamic properties, prediction of the future states can be made. The dynamic patterns in the financial market keep changing with respect to time.

Hence we dynamically choose the best time scale. Depending upon the real part of eigenvalues, the profit or loss of a company can be determined by its corresponding growing modes and decaying modes respectively.

The procedure involved in DMD algorithm is explained as follows:

At a particular instant,the data snapshot of the system in the matrix:

$$X_1 = \begin{bmatrix} | & | & | & & | \\ x_1 & x_2 & x_3 & \cdots & x_m \\ | & | & | & & | \end{bmatrix}$$

The delayed version of the above matrix is:

$$X_2 = \begin{bmatrix} | & | & | & & | \\ x_2 & x_3 & x_4 & \cdots & x_{m+1} \\ | & | & | & & | \end{bmatrix}$$

DMD finds a matrix A such a way that,

$X_2 = AX_1 ----> \text{eq}(1)$

i.e it maps:

$$\begin{matrix} | & & | & & | & & & & | & & | \\ x_1 & \rightarrow & x_2 & \rightarrow & x_3 & \rightarrow & \cdots & & x_m & \rightarrow & x_{m+1} \\ | & & | & & | & & & & | & & | \end{matrix}$$

To peform DMD for a given data,we will be following these steps:

Step 1:

SVD of X_1 gives,

$X_1 = U\sum V^*$

Step 2:

$A=X_2X_1^{\dagger}$

$A=X_2V\sum^{-1}U^*$

Now we introduce a low rank matrix which is equivalent of A.

$\tilde{A} = U^*AU=U^*X_2\sum^{-1}U^*U$

$=U^*X_2V\sum^{-1}$

Step 3:

$\tilde{A}\,W=W\Lambda$

(Here W is the eigen vectors and Λ is diagonal matrix of eigen values)

Step 4:

$\phi=X_2V\sum^{-1}W$

$X=\sum_{i=1}^{r}\phi_i e^{\omega_i t}b_i$

eigen values -> time dynamics ; eigen vector -> dynamic modes

4.2 Higher Order Dynamic Decomposition

HODMD is an extension of DMD, which relies on following Koopman assumption:

Koopman assumption:

$v_{k+1} = Rv_{k,}$ for k=1,2......,k-1

Where v_{k+1}

$v_{k+d} = R_1 v_k + R_2 v_{k+1} + ... + R_d v_{k=d-1}$ for k=1,...K-d,

which relates d subsequent snapshots.

when d=1, HODMD is equivalent to DMD.

4.2.1 Higher Order Singular Value Decomposition

Higher Order SVD is a robust extension of SVD. This method considers the input multi-dimensional data as tensors.HOSVD decomposes the tensor T as follows:

$$T_{i_1, i_2, ..., i_N} = \sum_{n_1=1}^{r_1'} \sum_{n_1=1} .... \sum_{n_N=1}^{r_1'} S_{n_1 n_1 ...... n_N} U^1_{i_1 n_1} U^2_{i_2 n_2} ....U^N_{i_N n_N}$$

such that $i_1 = 1, 2, ..., I_1$; $i_n = 1, 2, ..., I_N$

r- denotes the ranks of the parts of the tensor.

U- represents mode matrices.

S- represents core tensor

4.3 Mean Absolute Percentage Error.

The mean absolute percentage error (MAPE) is a measure of how accurate a forecast system is. It is determined as the average absolute percent error for each time period minus actual values divided by real values, and it is expressed as a percentage. MAPE is scale-independent and can successfully help in comparing the forecast performance between different time series. It is calculated by taking the mean or average of the absolute difference between the actual and predicted value, which is divided by the actual value.

$$M = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$$

V.Methodology :

Load the input csv file in python using Pandas library. While loading, we parse the dates column so that the format of all dates is uniform, i.e., YYYY-MM-DD. Then, Split the input time series data into train and test data. Use the train data to fit the models and test data to find the performance of model using its forecast.

Method #1:

Using a single company's stock data, apply the three models :

1. ARIMA
2. DMD
3. HODMD

Find the MAPE values for each based on short-term and long-term analysis. Short-term analysis means predicting the stock prices for the next 7 days while long-term analysis is the prediction for the complete test data.

Method #2:

Use the combination of companies from the same sectors and the combination between different sectors and apply DMD to compare the performance between inter-combination and intra-combination based on the MAPE.

While applying DMD and HODMD, we apply a window size on the train data which defines the size of the input matrix.

For example, Let the number of data points be M, and the window size was chosen to be N, then the input matrix size will be : **(M-N+1, N)**.

We vary the window size for both short-term analysis as well long-term analysis and try to find the best window size for each based on the least MAPE value observed.

VI.Results:

As mentioned before we have used two methods to forecast the value of stocks. We use this method to predict the stock prices of 3 different companies: IIFL, Tata Motors, and Sun Pharma. The results of single stock prediction are given in table 1. These results are when the forecast is done for a period of 49 days. The error for short-term prediction for a period of 7 days is also given in fig.1, fig.2, and fig.3 respectively.

	ARIMA(1,0,0)	DMD	HODMD
IIFL	8.694	7.448	5.4
Tata Motors	9.653	4.138	3.108
Sun Pharma	27.23	9.18	2.62

Table 1 Forecasting with single stock

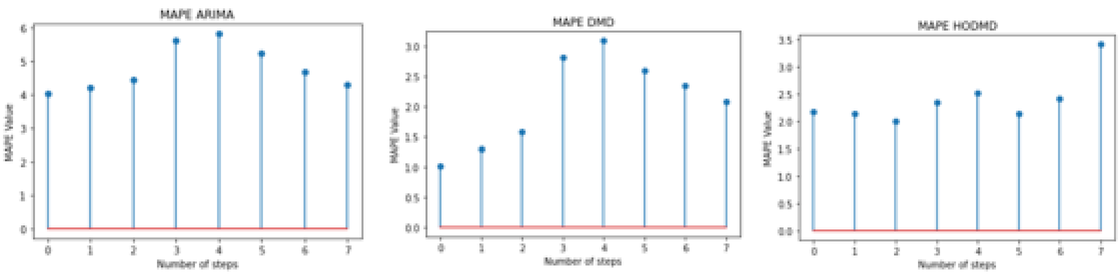


Fig. 1 MAPE values in short term for IIFL are given above. Window size used for DMD and HODMD was 5.

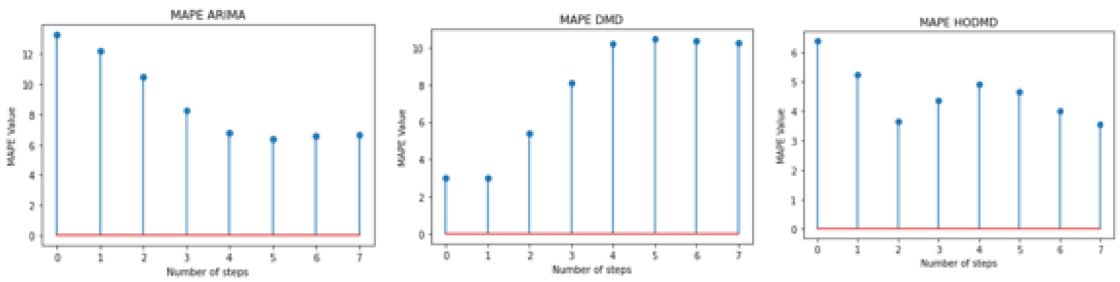


Fig. 2 MAPE values in short term for Tata are given above. Window size used for DMD was 10 and HODMD was 3.

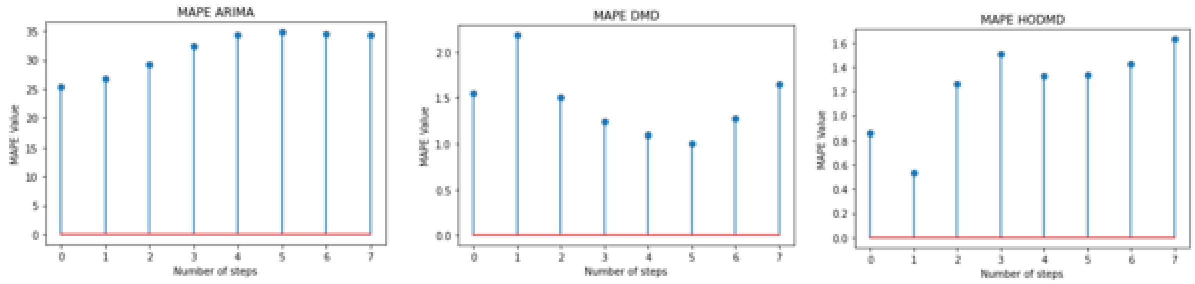
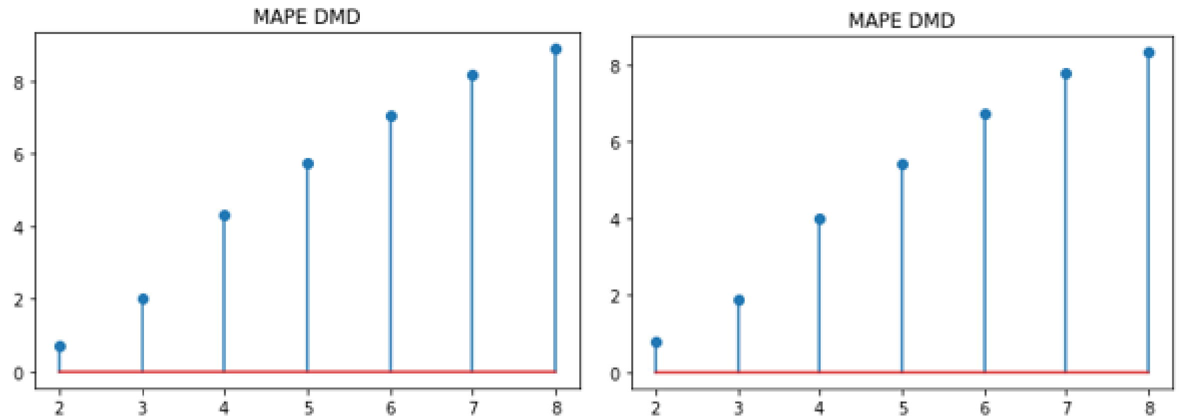


Fig. 3 MAPE values in short term for Sun Pharma are given above. Window size used for DMD and HODMD was 10.

From the table, it can be inferred that both DMD and HODMD outperform ARIMA. Out of the two HODMD has a better long-term prediction than DMD. It can be inferred from the figures that both DMD and HODMD perform better than ARIMA in the short term as well. In the short term, DMD outperforms HODMD but this does not happen in all cases.

In the second method, we combined stock prices from multiple companies from the same sector and from different sectors as well and then checked the performance of both. The company chosen for this is Sun Pharma. From fig. 4 it can be inferred that performance in the short term is similar when we take companies from the same sector or a combination of companies from related sectors.



In the long term it was observed that using companies from sectors related to the current sector gave a better result.

VII.Conclusion :

Overall it has been demonstrated that the method of HODMD (Higher Order Dynamic Mode Decomposition) performs better than the DMD and ARIMA in terms of long-term forecasting since HODMD got the least MAPE(Mean Average Percentage Error) value. In terms of short-term forecasting, there were instances where DMD outperformed the HODMD but still in most of the cases it was the latter that performed the best. Also, in the experiments that we have conducted where we used a combination of different companies belonging to the different sectors and a combination of different companies that belong to the same sectors so as to improve the forecasting of the stock price of a particular company, we observed that when a combination of different companies from different sectors are used to determine the stock price of a company it gave better results when compared to the combination of different companies from the same sector.

VIII.Future Works :

This work analyzes how an extension of DMD which is HODMD has outperformed its counterparts in stock price forecasting. Future works include using techniques such as Multi-Resolution DMD for our application and can be compared with the HODMD and other existing stock price forecasting techniques. Also, with HODMD we would like to explore its capabilities in other sectors, and we would also like to discover hotspots from the data using data-driven techniques.

IX.References

1. J. N. Kutz. Data-driven modeling and scientific computing: Methods for Integrating Dynamics of Complex Systems and Big Data. Oxford Press, 2013.
2. P. J. Schmid and J. Sesterhenn. Dynamic mode decomposition of numerical and experimental data. In 61st Annual Meeting of the APS Division of Fluid Dynamics. American Physical Society, November 2008.
3. Farmer, J. Doyne, et al. "A complex systems approach to constructing better models for managing financial markets and the economy." *The European Physical Journal Special Topics* 214.1 (2012): 295-324.
4. Grimm, Volker, et al. "Pattern-oriented modeling of agent-based complex systems: lessons from ecology." *science* 310.5750 (2005): 987-991.
5. Münnix, Michael C., et al. "Identifying states of a financial market." *Scientific reports* 2.1 (2012): 1-6.
6. Mann, Jordan, and J. Nathan Kutz. "Dynamic mode decomposition for financial trading strategies." *Quantitative Finance* 16.11 (2016): 1643-1655.

7. Kuttichira, Deepthi Praveenlal, et al. "Stock price prediction using dynamic mode decomposition." *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2017.
8. Cui, Ling-xiao, and Wen Long. "Trading strategy based on dynamic mode decomposition: Tested in Chinese stock market." *Physica A: Statistical Mechanics and its Applications* 461 (2016): 498-508.
9. Le Clainche, Soledad, and José M. Vega. "Higher order dynamic mode decomposition." *SIAM Journal on Applied Dynamical Systems* 16.2 (2017): 882-925.
10. Groun, Nourelhouda, et al. "Higher Order Dynamic Mode Decomposition: from Fluid Dynamics to Heart Disease Analysis." *arXiv preprint arXiv:2201.03030* (2022).