VIX Regime Clustering and Prediction based on Machine Learning

Ran Li

I. ABSTRACT

COBE Volatility(VIX)is a real-time market index representing the market's expectations for volatility over the coming 30 days and it is useful in making investment decision on stock market. In this article, we used Mixture of Gaussians to get regime information of the monthly VIX, predicted 1 time unit ahead VIX and discovered top forces that drive the regime switch behaviour in the stock market using monthly equity market volatility tracker(EMV) through OLS, ridge, LASSO and elastic net regression models. We will also present the direction for future studies in this article.

[Key Words]: VIX, Volatility Regime, Clustering, Gaussian Mixtures, Linear Regression, Elastic Net Regression

II. INTRODUCTION

VIX, short for CBOE Volatility Index, is a popular measure of the stock market's expectation of volatility based on SP 500 index options and investors usually use it to measure level of risk, fear, or stress in the market when making investment decision. Therefore, it is often referred to "fear index" [1]. Its derivatives also attracted a lot of attention thanks to its assistance in trading and hedging against changes in volatility[2]. Thus, predicting and clustering VIX has become useful for financial trading and analysis.

In this article, we are going to use Mixture of Gaussians to cluster monthly VIX index(from 1990/01 to 2021/10) into several regimes such that similar behaviours can be observed within each regime. We are also going to predict VIX and discover forces that drive volatility regime switch using monthly equity market volatility tracker(EMV) by constructing linear regression models, including ordinary least squares, ridge, LASSO and elastic net regression.

Note that EMV data is constructed by obtaining daily counts of articles containing at least one term in the categories economy or economic, uncertain or uncertainty, and one or more words from equity market, equity price, stock market and stock price. Over 1000 newspapers are used as samples and are retrieved from Access World News' NewsBank service [3]. In our study, we are going to use the monthly EMV with total 45 features including uncertainty from financial policy, government regulation to health care. And these are used as pool to trigger the volatility regime switch in our model.

We are going to introduce the models and optimization algorithm used in the methodology section. Results generated with Python are shown using tables and figures in the Results section and we put their interpretation, pros and cons of the models, future study direction in the Discussion section.

III. METHODOLOGY

Mixture of Gaussians (MoG)

In machine learning, we often encounter problem of clustering, which means we need to segregate groups with similar traits and assign them into clusters [4]. Mixture of Gaussians is a model usually used to achieve such goal. If we would like to find the distribution of a data point x_t, suppose our result contain k clusters, we can let C_j denote the j^{th} cluster, and use law of total probability to get the following formula:

$$f(x_t) = \sum_{j=1}^{k} P(C_j)P(x_t|C_j)$$

In this model, we assume that the probability of observing the data point if it comes to cluster j follows normal distribution with mean μ_j and standard deviation σ_j. Therefore, the probability density function should be

$$f(x_t) = \sum_{i=1}^{k} \pi_i (\frac{1}{\sqrt{2\pi_i}\sigma_i}) \exp \frac{-(x_t-\mu_i)^2}{2\sigma_i^2}$$

Since it is the linear combination of j normal random variables, we call this model Mixture of Gaussians. Based on given data points, we could find the optimal parameters, including all means, standard deviations and weights mentioned above. Let us use

$$\theta = \begin{pmatrix} \mu_1 & \sigma_1 & \pi_1 \\ \mu_2 & \sigma_2 & \pi_2 \\ ... & ... & ... \\ \mu_k & \sigma_k & \pi_k \end{pmatrix}$$

denote the parameter matrix. This can be done by minimizing the negative log likelihood function. Note that we have

$$
\begin{aligned}
\arg\min_\theta(-\log l(\theta|\hat{p})) &= \arg\min_\theta(-\log \prod_{j=1}^{M} p_j(\theta)^{n_j}) \\
&= \arg\min_\theta(-\log \prod_{j=1}^{M} p_j(\theta)^{\hat{p_j}N}) \\
&= \arg\min_\theta(-N \sum_{j=i}^{M} \hat{p_j} \log p_j(\theta)) \\
&= \arg\min_\theta(-N \sum_{j=1}^{M} \hat{p_j} log \frac{p_j(\theta)}{\hat{p_j}}) \\
&= \arg\min_\theta(-2N \sum_{j=1}^{M} \hat{p_j} log \frac{p_j(\theta)}{\hat{p_j}})
\end{aligned}
$$

where M is the number of bins we set when plotting the histogram of dataset, $\hat{p_j}$ is the observed proportion of points in bin j and $p(\theta)$ is the expected proportion on that bin, which can be computed by

$$
\begin{aligned}
P(\theta) &= P[b_j <= X <= b_{j+1}] \\
&= cdf_X(b_{j+1}) - cdf_X(b_j) \\
&= \int_{b_j}^{b_{j+1}} \sum_{i=1}^{k} \pi_i (\frac{1}{\sqrt{2\pi}\sigma_i}) \exp \frac{-(x-\mu_i)^2}{2\sigma_i^2} dx
\end{aligned}
$$

In this question, we first pre-processed data to obtain log return of VIX index by:

$$X_t = \log \frac{VIX_{t+1}}{VIX_t}$$

, and we aim to put X_t into k clusters, with the assumption that each cluster is generated from a normal distribution. Each cluster means a regime of stock market which measures volatility, how large swing of values around mean. In our experiment, we select $k = 2, 3, 4$ as number of regime to construct Mixture of Gaussians and it turns out that $k = 3$ will fit data best. Results and diagnostics for model fitting can be found in the next section.

Genetic Algorithm (GA)

Genetic Algorithm was developed by Jhon Holland in 1975[5]. We first randomly select some points in the searching space and evaluate the fitness of these points. While the termination criteria are not met, we can recombine, enlist, mutate or replace to obtain other solutions and evaluate their performance.

When we would like to build Gaussian Mixtures, we select Genetic Algorithm to perform the optimization. Specifically, we use it to find optimal θ by minimizing the negative log likelihood function above with constraints:

$$\sum_{i=1}^{k} \pi_i = 1, \pi_i \in [0,1], \mu_i \in (-\infty, +\infty), \sigma_i \in (0, \infty)$$

The constraint of sum for π_i may be troublesome because it prevents Genetic Algorithm's random selection among 0 and 1. Thus, we need to use an extra Lagrange Multiplier to help.The new objective function with LM included is written as

$$l(\theta) + \lambda|1 - \sum_{i=1}^{k} \pi_i| \ , \lambda > 0$$

Therefore, we need to optimize μ, σ and also this LM λ. We are going to discuss how we set the searching space(domain) of these parameters in practice using Python in Results section.

Linear,Ridge, Lasso and Elastic Net Regression

Regression models are built when discovering how uncertainty in various domains influence the volatility index in the next day where features ranging from politics, infectious diseases, macroeconomics to national security policy, etc. We build these regression models to not only predict the VIX in the next month but also to reveal leading forces that drive the volatility of the stock market by comparing their magnitude of regression coefficients. The regression models we used will be introduced as follows:

- *Linear Regression*
  We assume that the expected value of VIX in the next month is linearly dependent on the features in EMV in the previous month. Thus we have

  $$VIX_{t+1} = \mathbf{w^T X_t}$$

  And the estimates of coefficients w can be found by minimizing loss function using Ordinary Least Square

  $$argmin_w \sum_{i=1}^{N} (y_i - \mathbf{w^T X_i})^2 = \mathbf{X^T X}^{-1} X^T y$$

- *Ridge Regression* The central assumption of ridge regression is similar as linear regression except that we add an L^2 norm penalty on the loss function to decrease the risk of overfitting from regular linear regression. Thus we have

  $$argmin_w \sum_{i=1}^{N} (y_i - \mathbf{w^T X_i})^2 + \lambda ||\mathbf{w}||_2^2 = (\mathbf{X^T X} + \lambda \mathbf{I})^{-1} X^T y$$

  We know that if data matrix X has highly multicolinearity, we may face obstacle in inverting $X^T X$. However, after adding the hyperparameter λ, we need to inverse $\mathbf{X^T X} + \lambda \mathbf{I}^{-1}$ instead of the original matrix. Indeed, a small disturbance is added.

- *LASSO Regression*
  LASSO regression has the same idea as ridge regression and the only difference is that we use $L^1 norm$

penalty. Therefore we need to minimize the following:

$$argmin_w \sum_{i=1}^{N}(y_i - \mathbf{w^T X_i})^2 + \lambda \sum |w_j|$$

This can be solved using gradient descent and the result shows that some of the feature coefficients can be shrunk to 0, which means that LASSO regression has the power to perform variable selection. Those variables who have strong linearity with others will be removed from the regression model automatically.

- *Elastic Net Regression*
  Invented by Hastie in 2005 [8], elastic net regression is actually the combination between Ridge regression and LASSO regression, we use gradient descent to solve:

$$argmin_w \sum_{i=1}^{N}(y_i - \mathbf{w^T X_i})^2 + \lambda(\alpha \sum |w_j| + \frac{(1-\alpha)}{2} w_j^2)$$

  It helps resolve the problem that in LASSO regression, variable selection can be too dependent on data and thus unstable[6]. By taking into account some portion of ridge regression, we can switch between lasso and ridge regression. The coefficient α is l1-ratio and when $\alpha = 0$, this model is just ridge regression, when $\alpha = 1$, it is pure LASSO regression.

IV. RESULTS

Stock Market Regime Clustering

We first try to cluster the log return of VIX using Mixture of Gaussians. Let k denote the number of regime(cluster) of data points, we tried to build model for $k = 2, 3, 4$ respectively and estimates are obtained using the `geneticalgorithm` package in Python. However, it is worth noting that if we directly put the domains of parameters in the previous section, we are going to not only have very low efficiency of running but may also lead to wrong answers sometimes. For example, the sum of optimal π_js is greater than 1. Therefore, we should be very careful to determine the searching space of genetic algorithm. Table 1 summarizes the variable bounds used to generate result in this article and it is suggested that these bounds can be decided based on experience and observation. Here we choose $\mu \in [-1.0, 1.0]$ because the minimum and maximum of data points are -0.48 and 0.85. Similarly we can choose sigma between 1e-5 and 2.0. The Lagrange Multiplier should be comparably large because we want $|1 - \sum_{i=1}^{k} \pi_i|$ to be as small as possible. Lastly, in order to get sufficient convergence, we set max iteration in the algorithm as 1000 and the population size of initial generation as 100.

Also, in this article, we choose number of bins $M = 48$, that is because if our bin is so narrow, there is not a lot of data points fallen in that bin and if bins are too wide, data points concentrated on some certain bins and will lose distribution information to some extent. This M can be chosen by doing experiments.

See Figure 1 for the convergence of genetic algorithm. The convergence are similar for other cases as well. Table 2 shows the results of parameters for $k = 2, 3, 4$ and also their optimized likelihood function.

Model Diagnostics and Selection
In order to determine whether models that we built are valid and which k is best, we can test their goodness of fit using chi-squared test with degree of freedom $M - 3k - 1$. Results are shown in the last column of Table2. Note that p-value greater than 0.05 means that the model fits our data pretty well.

Therefore we find that $k = 2$ and $k = 3$ are good models but $k = 4$ will generate unacceptable model. In this article, we are going to choose $k = 3$ for further analysis. The reason is that $k = 3$ model has greater p-value and clustering the stock market into three regimes, each representing "low, medium and high" volatility will make more sense in analysing the behaviour than using only low and high volatility category.

Based on these estimates, we are able to find the regime that was primarily dominant at any given month using maximum conditional probability. If we define $S_t = i$ as the state of X_t at a given month t, then the probability that regime generating x_t is

$$P(S_t = i | x_t, \theta) = \frac{f_i(x_t)}{f(x_t)}$$

Obviously, the i that gets the largest value above would be the dominating regime at that month. Moreover, we could compute the transition model based on the dominating regime, that is, we can find the probability of regime switching from i to j using formula

$$\hat{p_{ij}} = \frac{number(X_t = i, X_{t+1} = j)}{N - 1}$$

Therefore, the transition matrix is

$$\begin{pmatrix} 0.20930233 & 0 & 0.79069767 \\ 0. & 0 & 1. \\ 0.10828025 & 0.0031847 & 0.88853503 \end{pmatrix}$$

where the first row represents medium volatility, second row represents high volatility regime and the last row represents the low volatility regime.

Driving Forces for Regime Switch, VIX Prediction

As discussed in the previous section that we would like to use the uncertainty information from EMV to

μ	-1.00	1.00
σ	0.00001	2.00
π	0.00010	0.99
λ	10000	100000

TABLE I: Variable Bounds in Genetic Algorithm

k	Coefficients (θ)	Likelihood	p value
2	$\begin{pmatrix} 0.128 & 0.260 & 0.171 \\ -0.025 & 0.149 & 0.828 \end{pmatrix}$	45.87	0.2773
3	$\begin{pmatrix} 0.0908 & 0.213 & 0.25 \\ 0.197 & 0.316 & 0.029 \\ -0.046 & 0.142 & 0.718 \end{pmatrix}$	44.62	0.3221
4	$\begin{pmatrix} 0.118 & 0.465 & 0.049 \\ -0.041 & 0.240 & 0.262 \\ 0.084 & 0.219 & 0.270 \\ -0.023 & 0.101 & 0.417 \end{pmatrix}$	57.12	0.04

TABLE II: Genetic Algorithm Results

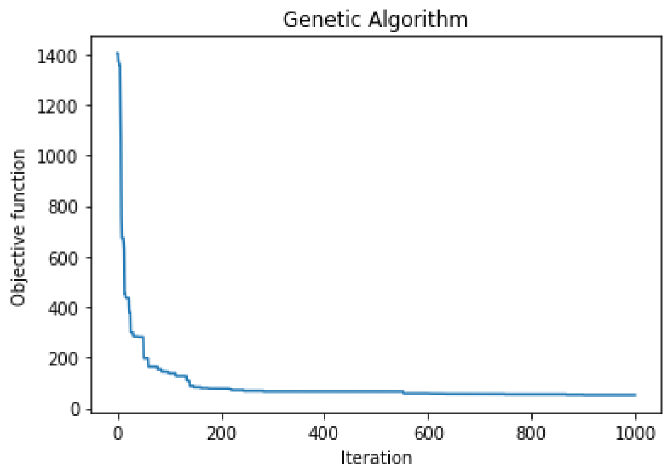

FIG. 1: k=3, Genetic Algorithm Convergence.

predict the VIX in the next month by constructing regression models. We built linear, ridge, LASSO and elastic net regression for the entire dataset. Note that since VIX indeed follows Markov process, which means that only information from previous time can influence the value for the next time unit [9]. This nice property actually makes cross validation applicable to choose best hyperparameters. Therefore, we randomly split 80% of data to train the model and use the remained 20% to test and compare the fitness of models and their mean squared errors. Results for these regression models can be found in Table3. We are going to interpret and discuss about the models in more details in the next section.

Since we also aim to discover what are the dominating forces that drive the change of volatility regime that is revealed in the previous question, we also construct Elastic Net Models separately on each segment of months with constant regimes. Note that using the dominating regime we found in the last section is doable but in this article, for simplicity, we use the following figure as guide to separate the whole dataset into 6 segments. For each segment, the regime changes significantly. After constructing models on each segment (hyperparameters are determined using cross validation), we sort the magnitude of the weights and find the top 5 forces that influence the change of VIX. Results are summarized in Table 4. More analysis and interpretation will be found in the discussion section.

Model	R^2	MSE	λ	L1 Ratio
OLS	0.651	4.99	-	-
Ridge	0.635	4.697	10.0	-
LASSO	0.59	4.66	0.16	-
Elastic Net	0.59	4.66	0.16	1

TABLE III: Regression Results Comparison

V. DISCUSSION

Let's first take a look at the result of Gaussian Mixture that we found in the last section using 3 as number of regimes in Table 2. Note that the first column of θ represents mean of that component, second column represents standard deviation and the last column represents its proportion in the market. Higher standard deviation implies higher volatility in the market. Therefore, we know that the second row the second regime whose $\sigma = 0.316$ is the regime representing highest volatility, the third row with σ 0.142 is the regime with lowest volatility. Combining with the ratio π, it is obvious to see that the market has very high volatility for only 2.9% of time and for almost 72% of the time, the market is relatively calm. This is not surprising because we rarely observe great fluctuations in stock, instead just mild swing.

Given the clustering result, we are able to construct the transition matrix, which tells us how regime switch given information in the previous month. Since the second row represents high volatility, we can find that for a day experiencing high volatility, in most of the time, it will switch into the state of low volatility. Similarly, we find that for all rows of this transition matrix, no matter which regime we start with, there is always greatest probability to switch into the low-volatile mode or stay in the low-volatile mode. This also shows that the stock market tends to be calm for most of the time.

In order to predict VIX, we considered EMV and tried to model VIX in time $t+1$ using EMV in time t. Table 3 shows result of regression models on the entire dataset. Among these 4 models, the regular regression model and Ridge regression model have highest R^2. That is

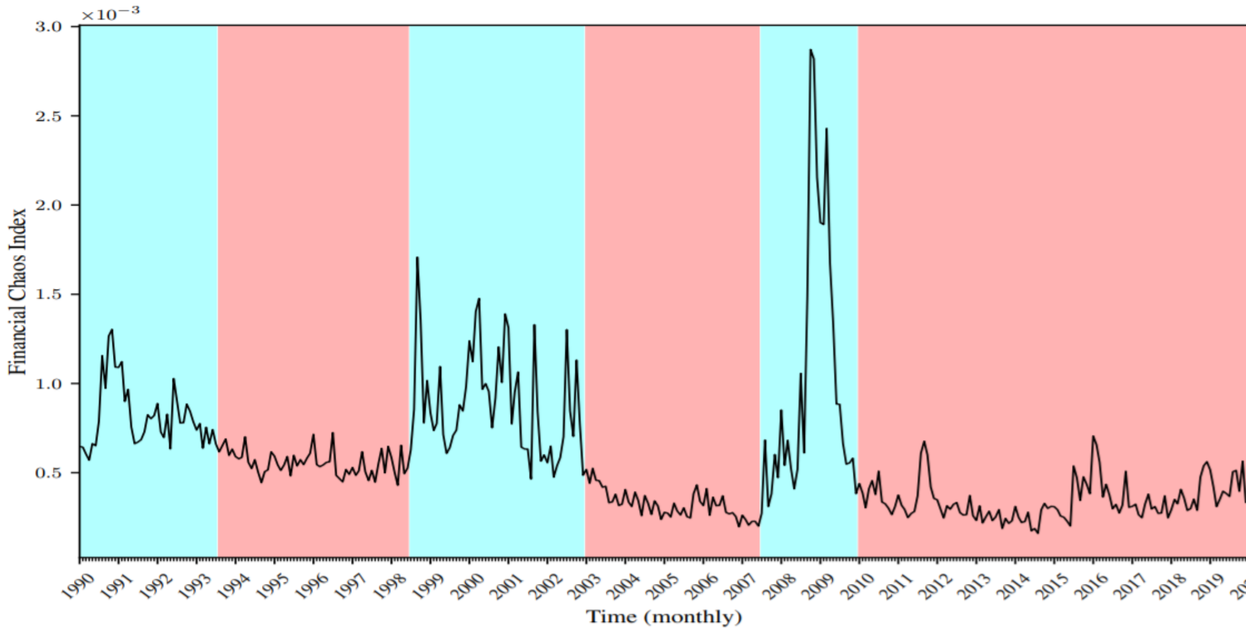

FIG. 2: Segments of VIX.

Date	R^2	λ	**l1-Ratio**	**Uncertainty 1**	**Uncertainty 2**	**Uncertainty 3**
1990/01 - 1993/06	0.824	0.06	0.1	Competition Policy (-1.802)	Macro – Real Estate Markets (1.8015)	Macro – Other Financial Indicators (1.506)
1993/07 - 1998/06	0.55	0.526	1.0	Macro – Interest Rates (-1.309)	EMV (1.0624)	Macro – Broad Quantity Indicators(-0.175)
1998/07 - 2002/12	0.269	9.27	0.1	Fiscal Policy (0.127)	National Security Policy (0.116)	Monetary Policy (-0.108)
2003/01 - 2007/09	0.656	0.24	1.0	Fiscal Policy (1.496)	Entitlement and Welfare Programs (-0.7820)	Macro – Labor Markets (-0.538)
2007/10 - 2009/12	0.83	0.979	0.1	Monetary Policy (-1.10)	Government-Sponsored Enterprises (1.023)	Macro – Consumer Sentiment (-0.756)
2010/01 - 2019/12	0.326	0.977	0.1	Macro – Real Estate Markets (0.345)	Financial Regulation (0.287)	Macro – Broad Quantity Indicators (-0.265)
1990/01 - 2019/12 (Entire Dataset)	0.596	0.16	1.0	Labor Disputes (2.387)	Financial Regulation (1.29)	Macro – Interest Rates (-1.01)

TABLE IV: Regression Results on 6 Segments

because in LASSO and elastic net regression, some of the features' weights may shrink to 0, thus we have less R^2. But luckily, it is still relatively efficient. On the other hand, we find that LASSO and Elastic Net regression have lowest Mean Squared Error(MSE), which means that these two models perform best in generalization. The reason why ordinary least square has the largest mean square error may be due to high multicolinearity among features and the other three models may help to decrease this issue. Note that the l1-ratio of elastic net regression is 1. If we consider the loss function of this model introduced in the Methodology section, it is obvious that the elastic net here is purely LASSO regression, which means that we have eliminated some features during the process. In addition, in the last row of Table 4, we find the top 3 leading uncertainty forces influencing the volatility over the past 50 years are "Labor Disputes", financial regulation and interest rate.

There are more insights if we move our concentration on elastic net regression on segments. As introduced, each segment seems to share same regime of volatility. Among these 6 segments from 1990/01 to 2019/12, the first, fourth and fifth model have $R^2 > 0.6$, meaning that these models explains our data well. If the l1-ratio is 1, then it means the features that influence volatility change are sparse and if l1-ratio is 0.1, then it means that there tend to be various weighted features influencing the market. For example in segment five, from 2007/10 to 2009/12, we can observe great fluctuation on the stock market. The result shows that during this time, l1-ratio is 0.1, which means that a lot of factors are weighted and it is their joint affect on the market that produced such big fluctuation. Among all these features, monetary policy, government-sponsored enterprises and consumer sentiment have largest impact to the VIX index. This makes sense in real life because the world was experiencing huge financial crisis caused by cheap credit and lax lending standards that fueled a housing bubble. In order to save economy, the bank and government took actions to regulate the market. For example, The Emergency Economic Stabilization Act provided $700 billion in bailout relief[7]. Uncertainty from new policies and regulations may be most significant factors that push the market to switch into high volatility mode.

However, we find that for the second, third and sixth segment, our model does not fit very well due to their small R^2. The reason may be that our assumption of linear relationship may over-simplify the real market. Also, probably there are other triggers that are not described in EMV data that push the switch change. Another reason may lie in small sample size. In this article, we use monthly VIX and monthly EMV. If we would like to train a more complex model like neural network, we need far more data points. In the next step, maybe we can use daily trading data to discover more complicated models to describe the market better.

Also, since we would like to focus on the motivation of regime switch, we can put more windows on each segment and focus on the intersection of regimes. For example, we can try to discover which weight of uncertainty metrics suddenly increases causes regime of volatility switch.

To put everything together, in this article, we used Mixture of Gaussians to cluster log return of VIX index and constructed transition matrix based on maximum conditional probability. In order to reveal the motivation behind each regime switch, we built regression models and it turns out that elastic net regression performs best on most of the segments. Even though fitness is not high on all segments, we can get meaningful and insightful information to infer the switch of regime. For further deeper and more accurate analysis, a larger sample size and narrower window may be required.

VI. REFERENCES

[1] Carr, P. (2017). Why is Vix a fear gauge? Risk and Decision Analysis, 6(2), 179–185. https://doi.org/10.3233/rda-170123

[2] Konstantinidi, E., amp; Skiadopoulos, G. (2011). Are vix futures prices predictable? an empirical investigation. International Journal of Forecasting, 27(2), 543–560. https://doi.org/10.1016/j.ijforecast.2009.11.004

[3] Baker, S., Bloom, N., Davis, S., amp; Kost, K. (2019). Policy News and stock market volatility. https://doi.org/10.3386/w25720

[4] Reynolds D. (2009) Gaussian Mixture Models. In: Li S.Z., Jain A. (eds) Encyclopedia of Biometrics. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-73003-5_196

[5] Alam, T. (2020). Genetic algorithm: Reviews, implementations, and applications. https://doi.org/10.22541/au.159164762.28487263

[6] Machine Learning Note-ridgeregression, lasso regression,elastic net regression csdn. Retrieved December 25, 2021, from https://www.cnblogs.com/dataanaly/p/12972435.html

[7] Ross, S. (2021, September 21). What major laws were created after the 2008 financial crisis? Investopedia. Retrieved December 25, 2021, from https://www.investopedia.com/ask/answers/063015/what-are-major-laws-acts-regulating-financial-institutions-were-created-response-2008-financial.asp

[8] Zou, H., amp; Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

[9] Introduction to finite markov chains. (2017). Markov Chains and Mixing Times, 2–20. https://doi.org/10.1090/mbk/107/01

VII. APPENDIX

All the code used to generate results in this article can be found in the following link:

https://github.com/ranli123/VIX-Regime-Clustering-and-Prediction-based-on-Machine-Learning.git