Journal of Information Systems Engineering and Management
2025, 10(12s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Modeling Stock Market Dynamics in the Philippines Through Symbolic Dynamical System

Liane Vina G. Ocampo[1], Rachelle Santos[2], Teotima Evangelista Gorres[3], Reynold Campo[4]
[1]lianeocampo@cvsu.edu.ph
Cavite State University, Philippines
[2]ramsantos@aims.edu.ph
Asian Institute of Maritime Studies (AIMS)
[3]teotimaabato@gmail.com
Agusan del sur State College of Agriculture and Technology, Philippines
[4]reynold.campo@pcu.edu.ph
Philippine Christian University, Philippines

ARTICLE INFO

Received: 24 Nov 2024

Revised: 01 Jan 2025

Accepted: 21 Jan 2025

ABSTRACT

The study employs the theory of symbolic dynamical system in modelling stock market in the Philippines. Precisely stock market modeling process is complex and dynamic. The main objective of this paper is to build a model using the following factors namely; peso dollar exchange rate, inflation rate, GDP, un employment rate. These factors are assumed to be the indicators of the

Philippine stock market. In this study we derived the mathematical model of each of those factors. The output of this paper can build a possible programming system or software for forecasting stock market predictions to be more accessible and more accurate since various machine learning approaches have been applied in stock market prediction.

Keywords: Modeling Stock Market, Dynamical System

Introduction

This study introduce a new method of modeling chaotic system like stock market. A symbolic dynamical system is a new concept applied to a complex system. Due to various stock market research existing (Mintarya, L., and Halim, J., et al (2023) scrutinizes 30 different studies focusing on various machine learning approaches and models includes neural networks and support vector machines.

Symbolic regression (SR), a type of machine learning, offers a promising alternative by evolving mathematical expressions that best fit the observed data without assuming predefined model structure (Schimidt & Lipson, 2009). SR combined with genetic programming (GP) is a well-established technique for generating the mathematical model that illustrate relationship within dataset (Takaki & Tomoyuki, 2022). SR with GP infers both linear and nonlinear model. Unlike in classical regression in which the model and the parameters are linear and follows a Gaussian distribution. SR use to estimate the symbolic space (function space) of a dynamical system while GP use to estimate the parameters of a symbolic dynamical system.

Genetic programming is an evolutionary computation technique that utilizes an evolutionary algorithm, making it particularly suited for handling large datasets. Genetic programs are developed from genetic algorithms, which are inspired by principles of evolutionary biology. During the evolutionary process, computer programs undergo mutations or iterations to potentially produce optimal models. In evolutionary biology, life evolves from simple to complex forms through natural selection, where individuals with higher "fitness" are more likely to survive and reproduce. These high-fitness individuals are retained and paired with other high-fitness individuals to produce offspring. The goal is for the offspring to achieve higher fitness values than their parents. The algorithm concludes when an offspring meets a predetermined fitness measure (Gorres, T., 2021).

The development of Symbolic Regression Program

The development of Symbolic Regression (SR) programs has a rich history rooted in symbolic dynamical systems (SDS). Hadamard first applied SDS in 1898 to complex systems, and in the 1950s, Von

Neumann described SDS as a powerful framework for artificial systems. Genetic programming models, utilizing machine learning algorithms, were developed to extend this idea, including the symbolic regression programs by Koza (1994), Schmidt and Lipson (2009), and Cohen, Beykal, and Bollas (2023). System dynamics theory, particularly coding theory or symbolic regression programs for symbolic dynamics, involves building a state space (function or model), identifying statistical properties (Lind & Marcus, 2021), and modeling sequences of infinite or bi-infinite symbolic shift spaces. These are measured in discrete time intervals with shifts or symbolic invariant mappings between spaces, forming pattern modeling-based codes in a symbolic system (Marcus & Rosenthal, 2013).

Investigating dynamic and complex systems—whether physical, biological, or ecological—requires significant effort to achieve meaningful results. As systems are inherently complex, researchers have developed system identification techniques such as symbolic regression (SR) through artificial intelligence and modern dynamical systems (genetic programming). In modeling and investigating real-world problems, symbolic regression algorithms based on genetic programming (GP) are used to identify iterated curves or functions. SR represents an abstract model space in topological systems for machine learning techniques (Schmidt & Lipson, 2009; Märtens, Kuipers, & Mieghem, 2017), which can be explained by physical systems through experimental data.

Computer systems are employed to elucidate system dynamics and complexity (chaoticity) through probability theory, measure theory, and ergodic processes. Symbolic regression, used for prediction models in real-world problems, fits the complexity or dynamics of these systems.

The development of modern dynamical systems and symbolic regression programs has been an active research area, with significant contributions from Koza (1992), Vladislavleva (2008), and Schmidt and Lipson (2009). They emphasize that genetic programming in symbolic regression does not assume any predefined model for the input-output data during the evolutionary process.

Symbolic regression involves inferring a non-linear and linear mathematical model in a symbolic analytical function $f: R^n \rightarrow R$, which fits two random variables X, and such that $y_n =$

$f\ (x_1, \ldots \ldots . . x_n)$ given $X_i = (x_1, \ldots \ldots . . x_n)$ and $Y_i = (y_1, \ldots . y_n)$ $i = 1, \ldots .$ input-output data

(Tenachi, , Ibata, and Diakogiannis. (2023)

As explained by Poli, Langdon, and McPhee (2008), SR is a computer program or model used to fit numerical data within a specific interval. De Franca and Aldeia (2021) clarified that SR is used to model input-output data, identify systems, make predictions about unobserved samples (state space), and determine a system's statistical properties by simultaneously estimating coefficients or parameters and function space.

Conversely, symbolic regression, introduced by Vladislavleva (2008) and Gorres-Abato and Tarepe (2014), can be expressed as:

$$\hat{f} = (X\ F\ C) \qquad (1)$$

This represents a measure-theoretic space over time. De Franca and Aldeia (2021) consider the equation $M(X\ F\ C)$as the Interaction-Transformation for Symbolic Regression, where X is a state (function) space in an input-output pair (X, Y) F is a set of functions (function operators and analytic functions), and Crepresents the weights or coefficients. For example, equation (1) can be represented by a simple function resulting from mutation (iteration) using an evolutionary algorithm, a genetic programming (GP) approach (Poli, Langdon, & McPhee, 2008), to optimize the search space for an interpretable expression such as:

$$G(x_i) = 2 + 4x_1 - 4\sin x_2{}^2 x_1 + 7cos\ {}^{x}{}_{x}{}^{3}4^4 - 3\log x_5$$

and can be represented as in (X, F, C)

$X = [[1,0], [1, 2], [4, -1], [1,0]]$

$F = [[sin, cos, log]]$

$C = [2, 4, -4, 7, -3]$

Aside from Koza and Poli (2005), Kadierdan, K., Nathan, K., and Brunton, S., (2020) and various other researchers have identified that SR programs in modern dynamical systems use simultaneous parameter and function space estimation (Kay & Mowbray et al., 2023; Austel & Dash et al., 2017) within symbolic dynamical systems through measure theory, random processes, and ergodic processes, optimizing iterations of function space (model) (De Franca, 2018; 2022; Koza, 1992; 1994).

Methods and Design

A symbolic regression program (genetic programming) is a stochastic process optimizing the desired prediction model and the relationship of input and output data. In a dynamical system, the Symbolic regression is a function that describes the relation of the input to output data and can be

written as:

$$f = (X, F, \mu)$$

Where any input variables, F basic functions operators (+, -, sin, cos, etc.) and μ set of measures

$(x_1, x_2, x_3, \ldots, x_n) = S \quad \rightarrow (input\ data)$

The task of symbolic regression with Genetic programming is to identify the target expression

(function) or model in symbolic form which estimates the values of a specified target variable based on the values of a set of input variables, that is;

$\hat{S} = \hat{f} = (X\ F\ W) \rightarrow output\ data$

The study made use of the descriptive methods of research using evolutionary algorithms particularly the genetic programming based symbolic regression. Data were obtained from the Macrotrends Data

Download, Bangkok Sentral ng Pilipinas, Philippine Stock Exchange Inc (PSE), www.psa.gov.ph/content/summary-inflation-report. The maximization can be quantitatively measured by:

$$Minimize: MSE = \frac{\sum_{i=1}^{n}(S_i - \hat{S}_i)^2}{n-1}$$

Where: $\hat{S}_i$= predicted value of S_i using four (4) parameters.

The parameters are:

The terminal set are:

X = peso dollar exchange rate

Y =inflation rate

Z =GDP

W = un employment

The function set used are: addition, subtraction, sine function, cosine function, and constant function

$$\hat{S}_i = {}_1X + w_2Y + w_3Z + w_4W \quad (2)$$

Where: w_1, w_2, w_3, and w_4 are weights to be determined by using Symbolic Regression via genetic programming (GP). The program can generate the GP search process and establish proposed model. The software used is the license version of DATAROBOT.

RESULTS AND DISCUSSION

The analytic solution to the model using the Symbolic regression with Genetic programming mathematically expressed by:

$$S^{\hat{}} = M(X\ F\ W)$$

The resulting model:

$$S^{\hat{}} = 26.42 + 0.07Z - (X - 52.243) - 4.47\,(X - 47.901) \qquad (3)$$

Equations (3) is the model of stock market in the Philippines with minimum mean squared error (MSE)= 0.067. From the four (4) parameters used in the study only two (2) parameters shown in the output model as the result of the run. This model described that inflation and un employment rate has no contribution to the model as in equation (3) above.

Further research is recommended for the improvement of the model. Any collaboration for creating a software for this particular study is very much welcome.

Conclusion

Based on the result, it identified drivers of stock market activities to be the peso-dollar exchange rate and GDP. The result suggests that, within this study, the sensitivities of the stock market in the Philippines relate more to changes in the peso-dollar exchange and GDP, rather than to inflation or unemployment. Excluding the last sets of variables could lead to an inference that they may not directly and significantly affect stock market movements at least within the time period and conditions given in this research.

Recommendation

Further refinement and cross-validation approach should also be applied to check the model's efficiency in prediction. The research could be extended easily to include more predictors, such as interest rates, inflows of foreign investment, or indices reflecting consumer-sentiment factors that influence the stock market.

Furthermore, Policymakers and financial regulators ought to take into account the identified critical factors—the peso-dollar exchange rate and GDP—when developing strategies aimed at stabilizing or stimulating the stock market. Gaining insight into these interconnections can facilitate informed decisionmaking that fosters economic growth.

REFERENCES

[1] Mintarya, L., Halim, J., Angie, C., Achmad, S., and Kurniawan, A., (2023).

[2] Machine learning approaches in stock market prediction: A systematic literature review, Procedia Computer Science, Volume 216, 2023, Pages 96-102,

[3] ISSN 1877-0509, https://doi.org/10.1016/j.procs.2022.12.115.

[4] (https://www.sciencedirect.com/science/article/pii/S1877050922021937)

[5] Schmidt, M., & Lipson, H. (2009). Distilling Free-Form Natural Laws from Experimental Data. Science, 324(5923), 81-85.

[6] Takaki, Katsushi, Miyao, Tomoyuki [2022]. Symbolic regression for the interpretation of quantitative structure-property relationships, Artificial Intelligence in the Life Sciences,Volume 2, ISSN 2667-3185,https://doi.org/10.1016/j.ailsci.2022.100046.

[7] Ramdhane. F.B., (2023). Symbolic dynamical systems in topological spaces defined via edit distances. Mathematics [math]. Aix Marseille Université (AMU), Marseille, FRA.; University of Sfax, Tunisia, 2023. English. ffNNT : 2023AIXM0186ff. fftel-04146898f

[8] Koza, J.R., (1992). Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge, MA, USA, 1992. [Koza, 1994] John R. Koza. Genetic Programming II: Automatic Discovery of Reusable Programs. MIT Press, Cambridge Massachusetts, May 1994.

[9] Koza, J.R., (1994). Genetic Programming II: Automatic Discovery of Reusable Programs. MIT Press, Cambridge Massachusetts, May 1994.

[10] Schmidt, M., and Lipson, H., (2009). Solving iterated functions using genetic programming. [PDF] from psu.edu. Book Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers

[11] Cohen, B., Beykal, B., and Bollas, G., (2023)."Dynamic System Identification from Scarce and Noisy Data Using Symbolic Regression," 2023 62nd IEEE Conference on Decision and Control (CDC), Singapore, Singapore, 2023, pp. 3670-3675, doi: 10.1109/CDC49753.2023.10383906. keywords:

[12] Lind, D., and Marcus, B., (2021). An Introduction to Symbolic Dynamics and Coding. 2nd Edition.

[13] Part of Cambridge Mathematical Library. https://www.cambridge.org/ph/universitypress/subjects/mathematics/differential-and-integralequations-dynamical-systems-and-co/introduction-symbolic-dynamics-and-coding-2ndedition?format=PB&isbn=9781108820288

[14] Marcus, B., and Rosenthal, J., (2013). Codes, Systems, and Graphical Models - The IMA Volumes in Mathematics and its Applications 123. https://www.waterstones.com/book/codes-systems-and-graphical-models/brian-marcus/joachimrosenthal/9781461265344

[15] Smith, R., (2013). Introduction to Supercritical Fluids. Richard Smith, ... Cor Peters, in Supercritical Fluid Science and Technology, 2013. https://www.sciencedirect.com/topics/chemistry/statisticalmechanics#:~:text=Statistical%20mechanics%20uses%20probability%20theory,the%20macroscopic% 20(bulk)%20properties

[16] Märtens, M., Kuipers, F., & Van Mieghem, P. (2017). Symbolic Regression on Network Properties. In J. McDermott, M. Castelli, L. Sekanina, I. Wolf, & P. García-Sánchez (Eds.), Genetic Programming:

[17] Proceedings - 20th European Conference, EuroGP 2017 (pp. 131-146). (Lecture Notes in Computer Science; Vol. 10196). Springer. https://doi.org/10.1007/978-3-319-55696-3_9 de Franca, F. O. (2022). Transformation-Interaction-Rational Representation for Symbolic Regression.

[18] arXiv:2205.06807v1 [cs.NE] 25 Apr 2022

[19] de Franca, F.O. (2018). A greedy search tree heuristic for symbolic regression, Information Sciences, Volumes 442â€“443, 2018, Pages 18-32, ISSN 0020-0255,

[20] https://doi.org/10.1016/j.ins.2018.02.040. ttps://doi.org/10.1016/j.ins.2018.02.040 0020-0255/© 2018 Elsevier Inc. All rights reserved

[21] Vladislavleva, E. (2008). Model-based problem solving through symbolic regression via pareto genetic programming. [Doctoral Thesis, Tilburg University]. CentER, Center for Economic Research.

[22] Schmidt, M., Lipson, H. (2009). Symbolic Regression of Implicit Equations. In: Riolo, R., O'Reilly, UM., McConaghy, T. (eds) Genetic Programming Theory and Practice VII. Genetic and Evolutionary Computation. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-1626-6_5

[23] Tenachi, W., Ibata, R., and Diakogiannis. F., (2023). Symbolic regression driven by dimensional analysis for the automated discovery of physical laws and constants of nature. Journées 2023 de la Société Française d'Astronomie & d'Astrophysique (SF2A)., Jun 2023, Strasbourg, France. pp.107- 108. ffhal-04325284f

[24] Poli, Langdon, and McPhee, (2008). A Field Guide toGenetic Programming.

[25] https://www.researchgate.net/publication/216301261_A_Field_Guide_to_Genetic_Programming

[26] Koza, J.R., Poli, R. (2005). Genetic Programming. In: Burke, E.K., Kendall, G. (eds) Search Methodologies. Springer, Boston, MA. https://doi.org/10.1007/0-387-28356-0_5

[27] Kadierdan, K., Nathan, K., and Brunton, S., (2020). 2020SINDy-PI: a robust algorithm for parallel implicit sparse identification of nonlinear dynamicsProc. R. Soc.

[28] A.47620200279202002 79http://doi.org/10.1098/rspa.2020.0279

[29] Kay, H., and Kay, S., and Mowbray, M., et al., (2023). Symbolic regression based interpretable datadriven soft-sensor for process quality control.

[30] https://www.sciencedirect.com/science/article/pii/B9780443152740502432

[31] Austel, V., Dash, S., Gunluk, O., Horesh, L., Liberti, L., Nannicini, G., and Schieber, B., (2017).

[32] Globally Optimal Symbolic Regression. ArXiv:1710.10720 [stat.ML] (or arXiv:1710.10720v3 [stat.ML] for this version) https://doi.org/10.48550/arXiv.1710.10720

[33] Gorres, T., (2021). Symbolic Regression via Genetic Programming; Philippines Population Prediction: 2010-2020 (May 2021). International Journal of Science and Research (IJSR) 2021, Available at

[34] SSRN: https://ssrn.com/abstract=3869624 Gorres-Abato, T., and Tarepe, D., (2014). Unpublished master's degree thesis.

[35] Philippine Peso per US Dollar Rate . Bangko Sentral ng Pilipinas
[36] https://www.psa.gov.ph/content/summary-inflation-report-consumer-price-index-2018100-may-2024
[37] Software: license version of DATAROBOT