Sains Malaysiana 54(7)(2025): 1859-1873
http://doi.org/10.17576/jsm-2025-5407-18

Comparing Various Methods of Forecasting Stock Index Prices in a Shock-Affected Market: Based on Data Covering the COVID-19 Pandemic

(Membandingkan Pelbagai Kaedah Ramalan Harga Indeks Saham dalam Pasaran Terjejas Kejutan: Berdasarkan Data Meliputi Pandemik COVID-19)

REVATHI GANESAN[1] & R. NUR-FIRYAL[1,2,*]

[1]*Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia*
[2]*Center for Modelling and Data Analysis (DELTA), Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia*

Received: 12 November 2024/Accepted: 28 April 2025

ABSTRACT

The recent years have been economically challenging, with financial markets worldwide facing turmoil. From late 2020 to mid-2022, global economies were heavily impacted by the COVID-19 pandemic due to the implementation of lockdowns and austere quarantine measures that crippled world trade. As the pandemic abated with increased vaccination rates, global economic growth was expected to return to normalcy. However, commodities (especially oil), exchange rates, and stocks remain greatly devalued, continuing to restrict financial progress. The primary goal is to identify the most effective models for predicting the impact of international trade during COVID-19 pandemic. Accurate stock index forecasting is crucial in such uncertain economic conditions. Using Malaysia, Indonesia, and Singapore as the research targets, this study compares time series linear regression (TSLR), Bayesian regression, and support vector regression (SVR) in predicting major stock indices during pre- and post-vaccination periods. The models are evaluated based on root mean square error (RMSE), mean absolute error (MAE), and adjusted R^2 to determine their effectiveness. Results show that Bayesian regression outperforms other models in the pre-vaccination period due to its ability to incorporate prior information, whereas SVR performs better in the post-vaccination period, capturing complex market dynamics more effectively. These findings suggest that Bayesian regression is particularly useful during high-uncertainty periods, while SVR is better suited for stable market conditions. This method should be utilized extensively in future research with other machine learning methods to enhance forecasting accuracy, while additional macroeconomic variables such as inflation, interest rates, and geopolitical factors should also be considered. Furthermore, the findings of this study, shows that by incorporating Bayesian regression and machine learning can provide valuable insights for policymakers, investors, and financial analysts in navigating financial risks during economic crises.

Keywords: Bayesian regression; COVID-19; SVR; TSLR; vaccination

ABSTRAK

Ekonomi global pada beberapa tahun kebelakangan telah menghadapi cabaran ekonomi yang kritikal. Ekonomi global menghadapi kesan yang teruk akibat pandemik COVID-19 disebabkan oleh pelaksanaan sekatan dan kuarantin yang ketat yang merencatkan perdagangan dunia dari penghujung Disember 2020 hingga pertengahan 2022. Pertumbuhan ekonomi global dijangka kembali pulih dengan peningkatan kadar vaksinasi semasa pandemik. Walau bagaimanapun, nilai komoditi (terutamanya minyak), kadar pertukaran mata wang dan saham kekal merosot, yang terus mengehadkan kemajuan ekonomi. Model regresi telah menunjukkan prestasi ramalan yang baik dalam meramalkan harga indeks saham, tetapi objektif utama adalah untuk mengenal pasti model yang paling berkesan untuk menghadapi krisis ekonomi ini. Dengan menggunakan indeks dari Malaysia, Indonesia dan Singapura sebagai sasaran kajian, kajian ini bertujuan untuk membandingkan ketepatan ramalan model regresi linear siri masa, regresi Bayesian dan regresi vektor sokongan (SVR) untuk tempoh pra- dan pasca-vaksinasi bagi menentukan prestasi terbaik berdasarkan punca min ralat kuasa dua (RMSE), ralat min kuasa dua (MAE) dan nilai terlaras R^2. Model regresi Bayes dan SVR didapati berprestasi baik kerana kedua-duanya membenarkan anggaran parameter berdasarkan maklumat terdahulu, manakala kaedah klasik hanya boleh menggunakan data sedia ada. Model regresi Bayes dikenal pasti sebagai model yang paling berkesan dalam kajian ini kerana ia menunjukkan prestasi yang baik dalam data ujian. Kaedah ini harus digunakan secara meluas dalam penyelidikan masa depan dengan kaedah pembelajaran mesin yang lain untuk meningkatkan ketepatan ramalan dan faktor makroekonomi lain seperti inflasi, kadar faedah serta faktor geopolitik juga boleh dikaji.

Kata kunci: COVID-19; regresi Bayes; regresi linear siri masa; regresi vektor sokongan; vaksinasi

INTRODUCTION

Throughout modern history, it has been observed that stock markets are indicators of economic well-being as sources of statistical information that measure market fluctuations. The performance of a particular market segment is represented by its indices, which reflect key performance indicators of the economy such as stock prices, availability of raw materials, public sentiment, and supply–demand dynamics (Singh 2025; Zhao 2021). In today's interconnected global economy, financial markets are highly sensitive to international trade and macroeconomic events (Nirmala & Pundareeka Vittala 2023; Zhang 2023). Commodity markets such as gold and oil are key examples of this global influence (Al-Mousawi & Al-Ghalibi 2023; Fatima, Sohail & Sajid 2023; Li et al. 2022).

The COVID-19 pandemic led to severe financial disruptions, with stock markets experiencing sharp declines due to uncertainty. For instance, the UK's Financial Times Stock Exchange saw its worst crash since 1987, while Japan's stock market dropped over 20% (Chen 2020; Zhang, Hu & Ji 2020). Investor sentiment and macroeconomic conditions, such as GDP growth and industrial productivity were significant drivers of stock price volatility during the crisis (Ji et al. 2024; Li et al. 2022). In Malaysia, daily growth in active COVID-19 cases strongly corresponded with stock market declines, particularly in commodity markets like metal and oil (Razali & Nur-Firyal 2021). Meanwhile, the travel and hospitality sectors faced sharp downturns, while technology and e-commerce industries remained resilient (Mgalla 2023).

COVID-19 vaccinations had mixed effects on financial markets. While developing countries generally saw positive market reactions post-vaccination, developed markets displayed more varied responses (Oanh 2022). India's Bombay Stock Exchange (BSE) rebounded strongly post-vaccination, whereas ASEAN markets remained sluggish (Behera, Pasayat & Behera 2022; Herlina et al. 2022). Market sentiment improved following mass vaccinations but fluctuated when vaccination rates declined, highlighting the pandemic's unique financial implications compared to past crises (Benashvili 2020; Serkan, Çömlekçi & Ali 2022). These effects suggest the need for adaptive economic policies to mitigate risks associated with future global health crises (Baig, Darukhanawalla & Khan 2020).

The main objective of this study was to gain a better understanding of the relationships between historical stock indices and commodity market effects on international trade for specific ASEAN countries based on the impact of COVID-19. Given the heightened market volatility during the crisis, identifying key financial drivers is essential for improving predictive modeling and informing policymakers on risk mitigation strategies (Zhang, Hu & Ji 2020).

Many studies have examined financial market volatility using regression analysis and statistical forecasting techniques. Traditional models, such as multiple linear regression, have been used to predict stock movements for major companies, including Apple, Amazon, and Google (Cen 2022). More recently, machine learning techniques have enhanced financial forecasting by capturing complex market trends more effectively (Aziz, Barawi & Shahiri 2022; Hu et al. 2022). Time series methods, originally used for epidemiological modeling, have also been applied to forecast crude oil price volatility, which correlates with major financial indices (Hassan, Jati & Abidin 2023).

This study focuses on three of the best regression model types used in economic research today: TSLR, Bayesian regression, and SVR. Bayesian linear regression uses a prior likelihood distribution that closely tracks the impact of posteriori measures (Zhao 2021). As such, its regression coefficients and disturbance variances are treated as random variables, which makes the method flexible and intuitive for inferencing. Thus, Bayesian models are highly reliable in assessing variables that have a significant impact on the output of processes such as financial stock market forecasting (Zhao 2021). SVR, a machine-learning regression method, has significantly improved financial market predictability, particularly when using optimized kernel parameters (Joseph, Mishra & Rabiu 2019; Rawat, Singh & Sahu 2023). Another study concluded that optimizing parameter values to guide support vector machine (SVM) kernel functions can enable it to outperform nearly all other models in terms of parametric sensitivity (Hu et al. 2022). The SVM-based predictability of financial indices is known to be highly accurate compared to other models in this study (Yunianto et al. 2023). By evaluating these methods across pre- and post-vaccination periods, this study seeks to determine the most effective forecasting approach for volatile financial conditions.

METHODOLOGY

DATA COLLECTION

Data were collected from various financial market sources, such as Yahoofinance.com and Investing.com. Investing.com provides market quotes, data on commodities, futures, options, and stocks. Yahoo Finance is a component of Yahoo Network which offers original material, press announcements, financial reports, stock quotes, cryptocurrency information, and data analysis related to finance. Hence, our analyses are based entirely on secondary data. The focus areas include the Kuala Lumpur Stock Exchange (KLSE) in Malaysia, the Jakarta Stock Exchange Composite Index (JKSE) in Indonesia, and the Straits Time Index (STI) in Singapore to see the impact on neighboring countries. Moreover, the size and the way these countries conducted the after impact is different thus the impact of COVID-19 can be seen based on different policies. This study utilizes the closing prices of the stock indices of Southeast Asian countries such as Malaysia, Singapore and Indonesia, including daily data on the closing prices of the stock indices KLSE, STI, JKSE, CSI1000, Nikkei 225, BSE30 and S&P500, closing prices of oil, gold and exchange rates. Specifically, the percentage change of closing prices of these variables is used in this study.

According to World Health Organization, COVID-19 started to spread to other areas in Asia in January 2020 and COVID-19 vaccinations in Southeast Asia started in early 2021. Thus, the period of this study spans from January 2, 2020 to February 28, 2022, using 465 sample observations from each index to see the impact before and after COVID-19 in more details. The period criteria were divided into pre- and post-vaccination groups to ascertain all three models' changes in predictability for each index. By analyzing these distinct periods, we aim to assess how stock market movements evolved in response to vaccination campaigns. While vaccination played a key role in stock market recovery, other macroeconomic factors may have contributed. To account for this, we analyze stock indices across multiple countries to identify broader trends beyond vaccination-driven effects. However, external factors such as fiscal policies, government stimulus measures, and global economic recovery efforts may have also played a role in stock market movements during these periods. These influences should be considered when interpreting the results to avoid over-attributing changes solely to vaccination effects (Yang 2025).

The COVID-19 pandemic has negatively impacted the stock market, with less stable fluctuations before vaccination (Tan et al. 2022). However, positive investor sentiment towards vaccination and continuous vaccinations led to changes in stock returns and a positive reaction to financial markets after the vaccination (Kucher, Kurov & Wolfe 2023). Therefore, this study examined the pre- and post-vaccination data to find out the reliability and effectiveness of the methods used and the factors selected at each stage of time.

Based on the works of Fatima, Sohail and Sajid (2023) and Hairuddin and Nur-Firyal (2022), our independent variables are commodity markets (crude oil, brent oil, and gold). These commodities are linked with Southeast Asean countries' stock indices during COVID-19 pandemic (Al-Mousawi & Al-Ghalibi 2023; Fatima, Sohail & Sajid 2023). Based on previous studies, we chose the five-day relative difference in percentage (RDP5) of stock index effects (Chen 2020) as dependent variables. International trade results in closely interconnected global economy and stronger bilateral ties between countries (Nirmala & Pundareeka Vittala 2023). Thus, stock indices near Southeast Asia (BSE 30 from India, Nikkei 225 Index from Japan, and China Southern [CSI 1000] from China) are used to monitor their effects on stock markets in Malaysia, Singapore, and Indonesia. Finally, because US markets are known to affect global markets, we use the S&P500 and the USD exchange rate to observe their effects on South Asian economies (Chen 2020; Wong 2022). While secondary data sources provide a broad and reliable dataset, they may also introduce certain biases, such as discrepancies in reporting standards, data revisions, or missing values (Baldwin et al. 2022). These limitations could impact model accuracy and should be acknowledged when drawing conclusions. Additionally, market-specific anomalies, such as country-specific economic policies, may influence stock market behavior, requiring careful interpretation of results to avoid misattributions (Hu 2025; Li & Lin 2024). To mitigate these biases, missing values were removed using a complete-case approach, and all data sources were cross-checked for consistency. Table 1 shows the description and calculations for RDP5 and percentage change.

MODEL SELECTION

This study compares the performance of Time Series Linear Regression (TSLR), Bayesian regression, and Support Vector Regression (SVR). Applying all three methods in the pre- and post-vaccination periods provided six case results that were compared after model convergence and quality diagnostics were run. Then, based on model adequacy, which is determined using error measures such as root mean squared error (RMSE), MAE and adjusted R^2 value, the best model is identified.

TSLR is commonly used for financial market predictions due to its interpretability and ability to model temporal dependencies (Bharath et al. 2023). However, it assumes a linear relationship between variables, which may not always hold in volatile markets. Bayesian regression

TABLE 1. Description and calculation of RDP5 and the percentage change

Variable	Description	Calculation
RDP5 (5-day relative difference in percentage)	5-day relative difference in percentage of the stock index	$\frac{p(t) - p(t-5)}{p(t-5)} * 100$
Percentage change	The difference between closing price on a trading day and previous trading day in percentage	$\frac{p(t) - p(t-1)}{p(t-1)} * 100$

offers advantages by incorporating prior distributions, making it more robust in uncertain conditions. This is particularly useful in the context of the COVID-19 pandemic, where historical trends alone may not provide an accurate forecast (Jiet et al. 2024). SVR, a machine learning-based approach, is capable of capturing complex, nonlinear relationships in financial data, making it suitable for periods of heightened volatility (Zouaghia, Kodia & Ben Said 2024). By comparing these three models, we aim to determine which approach provides the most accurate and reliable predictions given the market conditions during the pandemic.

TIME SERIES LINEAR REGRESSION

TSLR forecasts a selected time series, Y_t, assuming that it has a linear relationship with the independent variables. The time series multiple linear regression model is shown in Equation (1):

$$\mathrm{Y}_t = \beta_o + \beta_1 \mathrm{X}_{1,t} + \cdots + \beta_k \mathrm{X}_{k,t} + \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + \varepsilon_t \tag{1}$$

where Y_t is the change in percentage of the stock index price; β_0 is a constant; and β_i represents the independent variable estimate after accounting for the effects of all other independent variables, where $i = 1,\ldots,k$. $X_{i,t}$ are the independent variables, where $i = 1,\ldots, k$. $Y_{t\text{-}p}$ are the lagged variables, ε_t is the white noise error term and t represents daily data.

The assumptions considered for this model are as follows: There is a linear relationship between the forecast variable, Y_t and the predictor variables, $X_{i,t}$; The differences between an observed Y_t and its expected value are caused by random errors; All ε_t are uncorrelated with each $X_{i,t}$; All ε_t are independent and normally distributed, with zero means and equal variance.

BAYESIAN REGRESSION

The Bayesian approach originated from Bayes' theorem, which follows the principle of conditional probability. The conditional probability of an event depends on the occurrence of another event, and the probability in this case is based on new potentially relevant information. In this approach, the probability distribution function is related to linear regression. The response variables are not expected to be chosen from a single probability distribution; however, the distribution itself is utilized to determine the approximation of the response (Zhao 2021). The highest posterior probability for the model parameter is determined based on the input and output of training data. The initial sample of observations, X, which follows a distribution with the vector parameter, had the posterior probability distribution as shown in Equation (2):

$$P(\theta|X) = P(X|\theta) * P(\theta) \tag{2}$$

Moreover, the posterior distribution for the model parameter is based on prior information combined with current data. Priors can be classified as 'informative', which give specific details about the variables, or 'non-informative', which give minimal knowledge. Weak or non-informative priors play a minimal role in the posterior distribution and can only give partial information without complete scientific knowledge about the parameter. Nevertheless, this information can provide rational and practicable limits to the posterior distribution, given the assumption that all parameters in the parameter space are equally likely.

A linear regression model indicates how a response variable is linked to a linear function of one or more predictor variables. The regression model is specified as follows:

$$\mathrm{Y}_t = \beta_o + \beta_1 \mathrm{X}_{1,t} + \cdots + \beta_k \mathrm{X}_{k,t} + \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + \varepsilon_t \tag{3}$$

where β_0 is the intercept and β_i is the vector of K regression coefficients. ε_i is the error term assuming that $\varepsilon_i \sim$ Normal $(0, \sigma^2)$. From Equation (3), as ε is normally distributed, Y is also normally distributed. Assuming β and σ are known and x represents vector of predictor variables, the likelihood distribution for Y is specified as follows:

$$P(Y|\beta,\sigma) \sim N(\mathbf{X}\beta,\ \sigma^2) \tag{4}$$

We performed Bayesian regression analysis using R software's stan_glm function in rstanarm. The models were fitted using Markov chain Monte Carlo (MCMC) methods based on the variants of the least squares estimation and the assumption that the likelihood distribution is normal (Gaussian) with an identity link function. We used weak informative priors as the default prior to stabilize the computations (Muth, Oravecz & Gabry 2018). The model parameter was estimated over multiple iterations and chains as needed for convergence, and sampling quality diagnostics (Monte Carlo standard error and effective sample size) were applied. Sampling quality diagnostics were used to confirm convergence and assess the effective sample size. After the model is fitted using MCMC, and assessed using posterior predictive checks, the best model with the lowest RMSE and MAE was chosen.

SUPPORT VECTOR REGRESSION

The SVM is a supervised machine-learning model used for classification and regression analysis. In regression tasks, SVR aims to find a function that best maps input features to continuous output values. It does so by identifying a hyperplane that fits the training data while minimizing error. A kernel function is used to transform the data into a higher-dimensional space, enabling SVR to capture complex patterns without excessive computational costs.

To ensure the best performance, SVR hyperparameters were systematically tested using a grid search approach, which automatically compares multiple parameter combinations to find the most effective settings. Selecting an appropriate kernel function is crucial for SVR, as it determines how well the model captures underlying patterns. The radial basis function (RBF) kernel was chosen due to its ability to model nonlinear relationships (Hu et al. 2022). The key hyperparameters—cost (C) and epsilon (ε)—were optimized using grid search with cross-validation to balance model complexity and generalization. C controls the trade-off between minimizing error and maintaining a simple model, while ε defines the acceptable error margin within which predictions are not penalized. The grid search systematically tested C values from 1 to 100 and ε values from 0 to 1 (in increments of 0.1), ensuring optimal SVR performance across different stock indices and market conditions. Cross-validation was applied within the tuning process to enhance generalization and reduce overfitting.

Given $S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ as training data, $x_i \epsilon R^d$, $y_i \epsilon R$, where $i = 1, \ldots, n$, and n is the number of training data points. d is the dimensionality of the training data set, x_i is the input vector of the model, and y_i is the desired target. The SVM discriminant hyperplane based on the kernel is then defined as follows:

$$f(x) = w'.\varphi(x) + b \tag{5}$$

where w is the weight vector; b is the bias; and φ is the feature extractor (i.e., a nonlinear mapping function). Using SVR, $f(x)$, which has the greatest deviation from target y_i, is obtained for all training data at optimal w and b. These coefficients are determined by solving the following optimization problem:

$$\text{minimize } \frac{1}{2}\|w\|^2 \tag{6}$$

$$\text{subject to } \begin{cases} y_i - w^T x_i - b \\ w^T x_i + b - y_i \end{cases}$$

To ensure the feasibility of the optimization problem, soft margin slack variables ξ_i and ξ^*_i are introduced such that

$$\text{minimize } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{i}(\xi_i + \xi_i^*) \tag{7}$$

$$\text{subject to } \begin{cases} y_i - w^T x_i - b \le \tau + \xi_i \\ w^T x_i + b - y_i \le \tau + \xi_i^* \\ \xi_i \xi_i^* \ge 0 \\ C > 0 \end{cases}$$

where $\xi_i + \xi^*_i$ is the empirical risk, ½ $(||w||)^2$ is the structural risk, and C is the regularization factor that determines the trade-off between the two risk types. The proposed optimization problem is then solved using the Lagrange multipliers a_i and a_i^*. Consequently, the general form of the SVM-based regression function is given as follows:

$$f(x) = \sum_{i=1}^{i}(a_i - a_i^*)K(x_i\, x_j) + b_j \tag{8}$$

where $K(x_i,x_j)$ is the kernel function (Smola & Schölkopf 2004). Any function satisfying the Mercer condition can be used, including polynomial, radial basis function (RBF), and sigmoid kernels (Cortes & Vapnik 1995).

RESULTS AND DISCUSSION

DESCRIPTIVE ANALYSIS

The descriptive statistics and correlation estimates of the variables used in this analysis are listed in Table 2 and Figure 1, respectively. The three dependent exchange variables (KLSE, STI, and JKSE) had mean returns of less than 0.1, with STI having a negative value. The standard deviations approximately 1 which indicates data points are clustered around mean, showing that all indices were relatively stable. The three commodity market variables were crude oil, Brent oil, and gold. Notably, the price of crude oil had widespread economic impacts due to its role as a global commodity (Al-Mousawi & Al-Ghalibi 2023). Based on Table 2, the standard deviation of crude oil was at 16.605, and Brent oil at 3.259 where both were higher than that of the other variables, indicating that crude prices and Brent oil price are more spread out from their respective average prices, thus, were unstable during the study period. Meanwhile, gold had standard deviation of 1.202, suggesting relatively lower volatility compared to crude oil and Brent oil. Stock indices from other countries (CSI 1000, Nikkei 225, BSE 30, and S&P500) had mean values less than 0.1 and the standard deviation was consistent across all indices and around their respective average prices, which indicates stability. The currency exchange rates for Malaysia (USD/MYR), Indonesia (USD/IDR), and Singapore (USD/SGD) had standard deviations of less than 1 and closer to their mean value, further indicating stable prices. This suggests that exchange rates may serve as a better indicator of import and export price effects (Wong 2022).

Figure 1 presents a heatmap visualization to provide a clearer representation of the relationships between the variables used in this study. Nearly all values were between 0.1 and 0.5, indicating moderate to low correlation among independent variables, suggesting that multicollinearity is unlikely. The period and dependent variables for each data item used in this analysis are listed in Table 3. The data is split into training and testing sets, with 70% used for training and 30% for testing.

TABLE 2. Descriptive statistics

Variables	Mean	Median	Mode	Standard Deviation	Minimum	Maximum
KLSE	0.055	0.110	-0.020	1.720	-12.662	5.921
STI	-0.014	0.000	0.000	1.364	-7.930	10.690
JKSE	0.008	0.035	0.380	1.346	-6.580	10.190
Crude oil	-0.892	0.060	0.110	16.605	-305.970	37.660
Gold	0.071	0.130	-0.090	1.202	-4.930	5.950
Brent oil	0.001	0.280	0.540	3.259	-24.400	14.430
CSI 1000	-0.006	0.060	0.040	1.504	-8.750	4.130
Nikkei 225	0.016	-0.030	-0.030	1.438	-6.080	8.040
BSE 30	0.055	0.140	-0.130	1.629	-13.150	8.970
S&P500	0.014	0.120	0.180	1.667	-11.980	9.380
MYR	0.005	0.000	0.000	0.280	-1.130	1.260
SGD	0.002	0.030	0.050	0.277	-1.170	1.160
IDR	-0.021	0.000	0.000	0.468	-4.410	1.910
RDP5	0.046	-0.234	0.235	2.509	-15.269	8.507

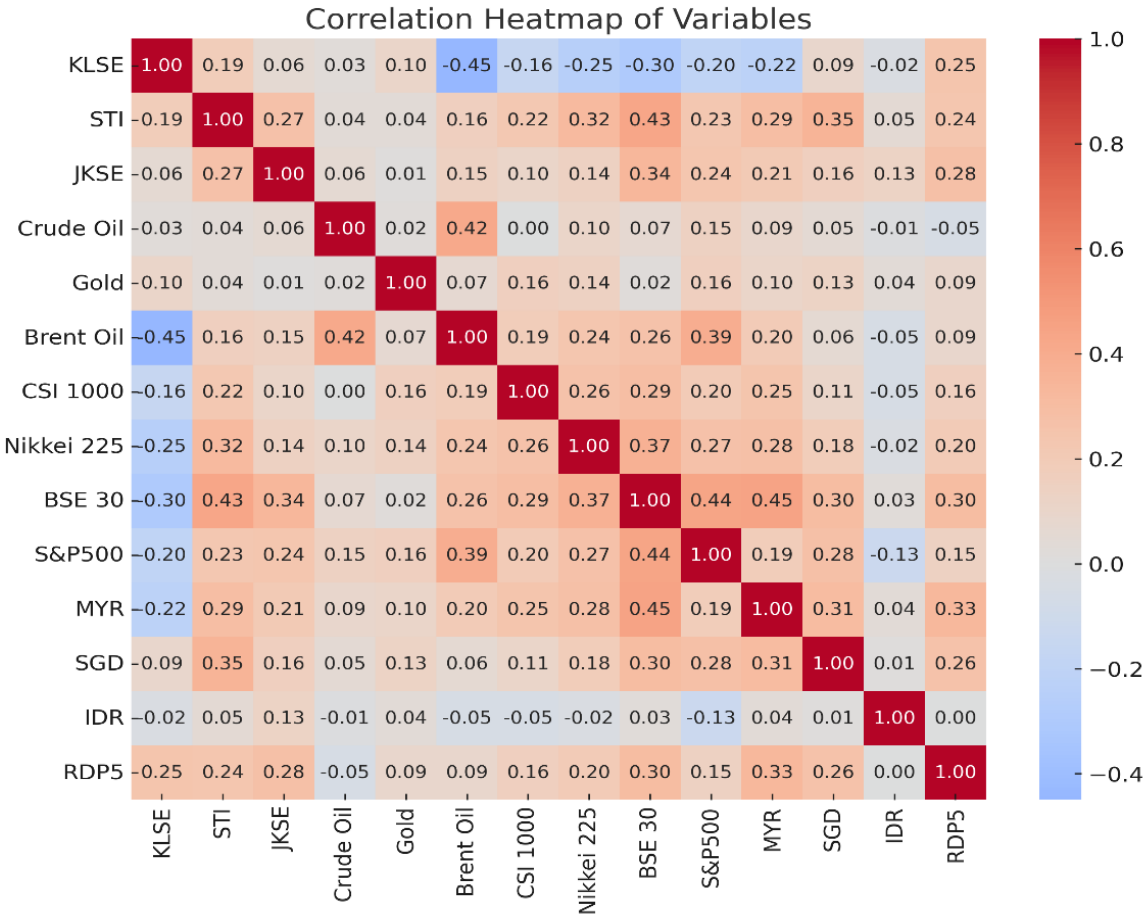


FIGURE 1. Correlation between variables

TABLE 3. Timeline of data

Data	Training data	Testing data
KLSE (Pre-vaccination)	161 (09.01.2020 – 26.10.2020)	69 (27.10.2020 – 26.02.2021)
KLSE (Post-vaccination)	135 (08.03.2021 – 08.11.2021)	58 (09.11.2021 – 22.02.2022)
STI (Pre-vaccination)	168 (09.01.2020 – 23.10.2020)	73 (26.01.2020 – 26.02.2021)
STI (Post-vaccination)	142 (08.03.2021 – 02.11.2021)	62 (04.11.2021 – 22.02.2022)
JKSE (Pre-vaccination)	156 (09.01.2020 – 21.10.2020)	68 (22.10.2020 – 26.02.2021)
JKSE (Post-vaccination)	137 (08.03.2021 – 09.11.2021)	59 (10.11.2021 – 22.02.2022)

TIME SERIES LINEAR REGRESSION

TSLR was conducted for all six variables. We checked the autocorrelation of the residual using the Breusch–Godfrey test and the ACF plots, then included lagged variables to improve model efficacy. Exchange rates were considered based on their respective stock index: where KLSE used MYR/USD, STI used SGD/USD, and JKSE used IDR/USD, while all other independent variables were included for all models. Table 4 shows the summary of significant variables of each model from the six different datasets. The full regression results, including coefficients (β values), standard errors, and p-values, are provided in Appendix A.

The results indicate RDP5 was the only consistently significant variable in all models, confirming its critical role in stock market movements (Chen 2020). Besides, the percentage change in the BSE Sensex 30 index was significant during the pre-vaccination period for all three stock indices, while Nikkei 225 index price was significant, especially for STI in pre- and post-vaccination periods. This echoes the findings of Nirmala and Pundareeka Vittala (2023), who found that Japanese and Indian markets influence broader Asian market trends. The exchange rate was only significant for JKSE in the post-vaccination period, suggesting delayed currency market adjustments. Variance inflation factor tests showed no multicollinearity among the variables.

BAYESIAN REGRESSION

With Bayesian regression, the same significant variables from the time series model in Table 4 were used. For this model, non-informative priors were used as the default prior function in R, assuming Gaussian distributions. The chains were set at values of four and 5000 × 2 iterations for all models, resulting in 20,000 posterior samples. Bayesian regression results, including posterior means, standard deviations, and 95% credible intervals, are provided in the Appendix A. Figure 2 illustrates the posterior predictive checks of all six models. The models provide reasonable posterior distributions, with the exception of STI and JKSE pre-vaccination, which show slight deviations, though still acceptable.

SUPPORT VECTOR REGRESSION

SVR leveraged the same significant independent variables listed in Table 4. Multiple kernel types were tested, and the radial basis function (RBF) kernel yielded the best results, consistent with prior studies on SVR-based stock price forecasting (Rawat, Singh & Sahu 2023). Hyperparameter tuning involved grid search with cross-validation to optimize C, gamma (γ), and ε, ensuring model generalization (Hu et al. 2022). Table 5 includes the optimal values for these parameters, improving model transparency. As described in the methodology, the support vectors are the closest points to the hyperplane obtained for each model. Notably, almost more than half the data points for each model in Table 4 were used as support vectors, indicating that most of data points were close to their respective hyperplanes. The similarity in hyperparameter values suggests that the market behaviors across KLSE, STI, and JKSE were comparable, requiring similar regularization and kernel settings. The consistent selection of the RBF kernel across models highlights its effectiveness in capturing nonlinear stock market patterns. If market-specific differences had been more pronounced, individualized hyperparameters for each index might have been necessary. However, as the selected values provided consistent and stable results across all datasets, they were deemed appropriate for this study.

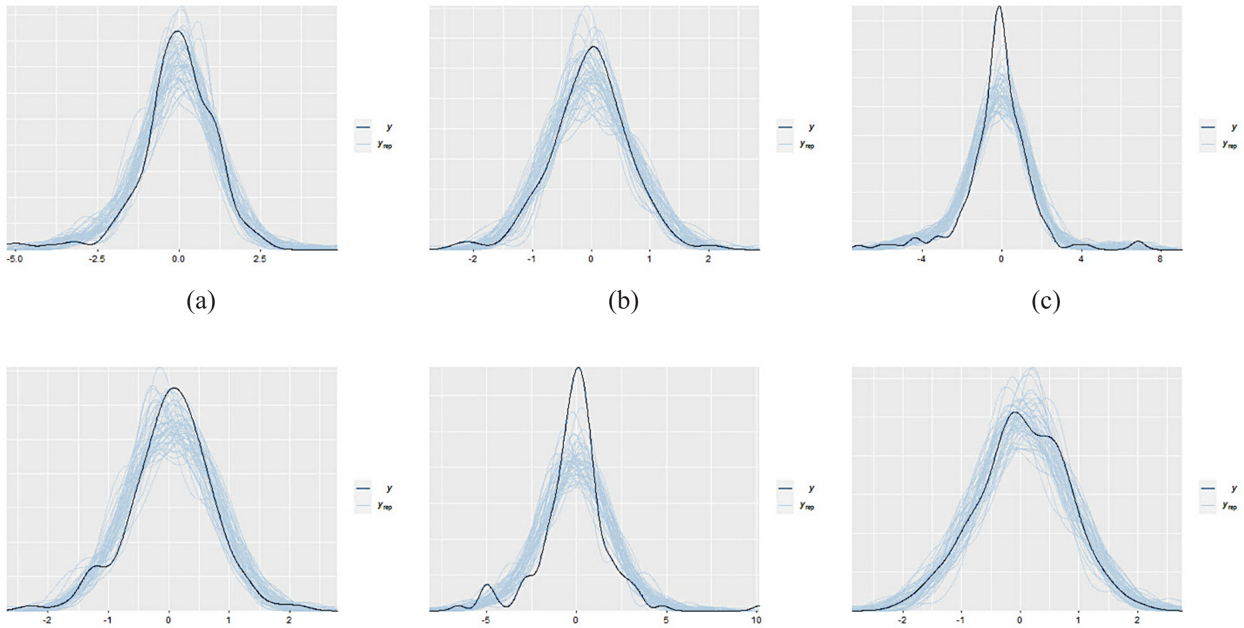
MODEL SELECTION

Table 6 compares the R^2, RMSE, and MAE values for all models and methods. The TSLR models for KLSE and STI pre-vaccination had R^2 values exceeding 0.5, indicating moderate explanatory power. Bayesian regression performed best in the pre-vaccination period, likely due to its ability to incorporate prior uncertainty and handle volatile markets. By treating parameters as random variables, Bayesian regression enables more stable model parameter estimation (Zhao 2021).

Unlike Bayesian and time series regression, which provide coefficient estimates and statistical significance tests like p-values, SVR focuses purely on predictive accuracy. As a machine learning model, it does not estimate

TABLE 4. Significant Variables in each model

Model	Significant independent variables
KLSE (Pre-vaccination)	Lag 2, RDP5, BSE 30, Nikkei 225
KLSE (Post-vaccination)	Lag 1, Lag 3, RDP5
STI (Pre-vaccination)	Lag 1, Lag 2, RDP5, BSE 30, Nikkei 225
STI (Post-vaccination)	Lag 1, Lag 2, Lag3, Lag4, RDP5, Nikkei 225
JKSE (Pre-vaccination)	Lag 2, Lag 6, RDP5, BSE 30
JKSE (Post-vaccination)	Lag 2, IDR, Nikkei 225, RDP5



(a) (b) (c)

(d) (e) (f)

FIGURE 2. Posterior predictive check for (a) KLSE (Pre-vaccination) (b) KLSE (Post-vaccination) (c) STI (Pre-vaccination), (d) STI (Post-vaccination) (e) JKSE (Pre-vaccination), and (f) JKSE (Post-vaccination)

individual predictor significance but instead minimizes errors within a specified margin. Therefore, SVR is evaluated using performance metrics such as R^2, RMSE, and MAE rather than hypothesis testing (Hu et al. 2022; Smola & Schölkopf 2004). In this study, SVR performed better post-vaccination, benefiting from its flexibility in modeling smooth, nonlinear trends.

Although SVR had the highest R^2 and lowest RMSE values, its MAE was higher, suggesting that it minimizes squared errors rather than absolute deviations, which may amplify large errors. This suggests that SVR might be more sensitive to outliers compared to Bayesian regression. The results indicate that adding more significant economic indicators or investor sentiment variables could further improve predictive performance across pre- and post-vaccination periods. To further enhance interpretability, visual performance comparison graphs (R^2, RMSE, and MAE plots) have been included in the Appendix A.

After analyzing the training data, the models were checked using testing data. Table 7 presents the MAE values for all six models. Empirical research suggests that using 20-30% of the data for testing purposes and the remaining 70-80% for training produces the most reliable results (Gholamy, Kreinovich & Kosheleva 2018). Thus, 30% of the original dataset was allocated for testing, as outlined in Table 3. Overall, Bayesian regression exhibited lower MAE values, indicating superior performance compared to TSLR and SVR. Notably, Bayesian regression

TABLE 5. Optimal parameters used for each dataset

Variable	KLSE Pre-Pandemic	KLSE Post-Pandemic	STI Pre-Pandemic	STI Post-Pandemic	JKSE Pre-Pandemic	JKSE Post-Pandemic
Lag 1	-	0.037	0.048	0.040	-	-
Lag 2	0.042	-	0.062	0.056	0.049	0.041
Lag 3	-	0.055	-	0.065	-	-
Lag 4	-	-	-	0.038	-	-
Lag 6	-	-	-	-	0.064	-
RDP5	0.061	0.073	0.073	0.046	0.078	0.045
BSE 30	0.079	-	0.042	-	0.047	-
Nikkei 225	0.052	-	0.047	0.050	-	0.072
IDR	-	-	-	-	-	0.057
Intercept	0.4260	0.1178	0.4793	-0.0838	0.3072	0.3482
Support Vectors	147	120	140	132	126	130
C	1	1	1	1	1	1
γ	0.1429	0.125	0.2	0.125	0.1429	0.125
ε	0.1	0.1	0.1	0.1	0.1	0.1

TABLE 6. R^2, RMSE and MAE values for all models

Data	Model efficacy	TSLR	Bayesian regression	SVR
KLSE (Pre-vaccination)	R^2	0.5288	0.552	0.5419
	RMSE	0.7947	0.785	0.795
	MAE	0.6296	0.5172	0.5468
KLSE (Post-vaccination)	R^2	0.4069	0.449	0.456
	RMSE	0.4988	0.485	0.4613
	MAE	0.3946	0.3219	0.3328
STI (Pre-vaccination)	R^2	0.7016	0.741	0.7012
	RMSE	0.8692	0.854	1.0455
	MAE	0.6388	0.5133	0.6187
STI (Post-vaccination)	R^2	0.4606	0.522	0.5556
	RMSE	0.4885	0.468	0.4475
	MAE	0.3566	0.2226	0.2982
JKSE (Pre-vaccination)	R^2	0.4667	0.583	0.5049
	RMSE	1.3956	1.24	1.3548
	MAE	0.9787	0.6258	0.7854
JKSE (Post-vaccination)	R^2	0.3595	0.412	0.4285
	RMSE	0.5998	0.585	0.5528
	MAE	0.4767	0.3966	0.4083

TABLE 7. MAE values for test data

Data	TSLR	Bayesian regression	SVR
KLSE (Pre-vaccination)	0.759	0.5161	0.7441
KLSE (Post-vaccination)	0.35	0.3872	0.3506
STI (Pre-vaccination)	0.7977	0.4456	0.7569
STI (Post-vaccination)	0.3553	0.2558	0.3755
JKSE (Pre-vaccination)	0.6622	0.6533	0.7226
JKSE (Post-vaccination)	0.4863	0.4444	0.5058

outperformed other models in the pre-vaccination period, reinforcing its ability to incorporate prior distributions and handle market uncertainty during volatile conditions. Conversely, SVR showed competitive results in post-vaccination datasets, suggesting that its flexibility in capturing nonlinear relationships made it well-suited for more stabilized market trends.

The findings of this study provide valuable insights for policymakers, investors, and financial analysts. Investors should consider Bayesian regression models in volatile markets due to their ability to incorporate prior distributions and handle uncertainty (Zhao 2021). In contrast, SVR models are better suited for stable market conditions, capturing nonlinear relationships effectively (Rawat, Singh & Sahu 2023). Regulators can use these insights to refine financial policies, ensuring model selection aligns with economic stability. Portfolio managers may also benefit from hybrid approaches integrating Bayesian priors with SVR's flexibility, improving financial predictions and risk assessments (Hu et al. 2022).

Despite these promising results, some limitations must be noted. The reliance on historical data means models may struggle with sudden economic shifts. Excluding sentiment analysis and high-frequency trading data limits capturing real-time investor behavior, which significantly impacts stock markets (Chen, Guo & Song 2024). Future studies should incorporate alternative data sources, such as news sentiment, macroeconomic trends, and social media analytics, to refine forecasting accuracy (Singh 2025). Additionally, while regression-based models offer strong interpretability, they may not fully capture complex nonlinear dependencies in financial time series. Future research should explore deep learning approaches, such as LSTMs, which can model long-term dependencies more effectively (Kabir et al. 2025). Hybrid models integrating Bayesian inference with deep learning architectures may further enhance predictive performance (Smola & Schölkopf 2004).

CONCLUSIONS

Developing stock index price prediction models is a challenging task, particularly during economic crises. This study applied TSLR, Bayesian regression, and SVR to forecast stock index prices before and after COVID-19 vaccinations, evaluating their predictive accuracy. The results indicate that Bayesian regression and SVR outperformed time series regression, with Bayesian regression excelling in testing data due to its ability to incorporate uncertainty and prior information. Additionally, all three models performed better with pre-vaccination data, suggesting that stock indices were more predictable before economic recovery. This finding aligns with research indicating that the relationship between independent variables and stock indices shifted post-pandemic (Tabash et al. 2022).

Given the limited studies on financial forecasting models in pandemic-induced crises, future research should focus on identifying key economic factors affecting stock index prices and optimizing model parameters for improved accuracy. Exploring advanced machine learning approaches, such as deep learning models, could enhance predictive performance. This study contributes to stock index forecasting in Southeast Asian markets, helping investors, financial analysts, and policymakers mitigate risks and enhance decision-making by understanding market fluctuations and recovery patterns.

REFERENCES

Al-Mousawi, Z.F.S. & Al-Ghalibi, K.S.H. 2023. Analysis of the relationship between crude oil prices and economic performance indicators in Iraq for the period 2004-2021. *Russian Law Journal* XI(12s).

Aziz, M.I.A., Barawi, M.H. & Shahiri, H. 2022. Is facebook PROPHET superior than hybrid ARIMA model to forecast crude oil price? *Sains Malaysiana* 51(8): 2633-2643.

Baig, U., Darukhanawalla, B.H. & Khan, Z.S. 2020. Deal with the fallout economic disorder in South Asian countries from COVID-19. *Journal of Social Sciences and Humanities* 59(2): 117-134.

Baldwin, J.R., Pingault, J.B., Schoeler, T., Sallis, H.M. & Munafò, M.R. 2022. Protecting against researcher bias in secondary data analysis: Challenges and potential solutions. *European Journal of Epidemiology* 37(1): 1-10.

Behera, J., Pasayat, A.K. & Behera, H. 2022. COVID-19 vaccination effect on stock market and death rate in India. *Asia-Pacific Financial Markets* 29(4): 651-673.

Benashvili, G. 2020. Pandemic impacts on the international economy. *Globalization & Business* 10: 192-197.

Bharath, K., Bhandari, G., Mehta, S. & Gupta, S. 2023. Commodity and stock price prediction using ML time series regression, LSTM. *International Journal of Advanced Research in Science, Communication and Technology* 3(1): 270-272.

Cen, X. 2022. Prediction on the price of technology stock and COVID-19 based on multiple linear regression. *BCP Business & Management* 23: 910-915.

Chen, S. 2020. Forecasting daily stock market return with multiple linear regression. *Mathematics Senior Capstone Papers*. 19. https://digitalcommons.latech.edu/mathematics-senior-capstone-papers/19

Chen, X., Guo, Y. & Song, Y. 2024. Multiple time scales investor sentiment impact the stock market index fluctuation: From margin trading business perspective. *The North American Journal of Economics and Finance* 69(PA): 102027.

Cortes, C. & Vapnik, V. 1995. Support-vector networks. *Machine Learning* 20(3): 273-297.

Fatima, S., Sohail, F. & Sajid, Y. 2023. An examination of impact of gold and oil prices on the stock markets indices during COVID-19. *International Journal of Financial Engineering* https://doi.org/10.1142/S2424786323500111

Gholamy, A., Kreinovich, V. & Kosheleva, O. 2018. Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation. *International Journal of Financial Engineering* 11(2): 105-111.

Hairuddin, H.B.M. & Nur-Firyal, R. 2022. Relationship between oil, gold and Malaysian stock market during COVID-19 movement control order using quantile regression and time series. *Sains Malaysiana* 51(7): 2315-2327.

Hassan, S., Jati, M.K.K. & Abidin, N.Z. 2023. The impact of COVID-19 on the Malaysia stock market: Finance sector. *International Journal of Advanced Research in Economics and Finance* 4(4): 120-127.

Herlina, M., Mafruhat, A.Y., Kurniati, E., Wildan, W. & Salsabila, H.G. 2022. The stock market reaction to COVID-19 vaccination in ASEAN. *F1000Res.* 11: 363.

Hu, H. 2025. A global comparative study of financial market anomalies. *Proceedings of the 3rd International Conference on Financial Technology and Business Analysis.* DOI: 10.54254/2754-1169/153/2024.19468

Hu, M., Tang, Z., Xie, X. & Jiang, M. 2022. Stock prediction and analysis based on support vector machine. *Frontiers in Business, Economics and Management* 5(2): 98-101.

Ji, X., Bu, N., Zheng, C., Xiao, H., Liu, C., Chen, X. & Wang, K. 2024. Stock market reaction to the COVID-19 pandemic: An event study. *Portuguese Economic Journal* 23(1): 167-186.

Jiet, M.M., Verma, P., Kamble, A. & Puri, C. 2024. A review on Bayesian methods for uncertainty quantification in machine learning models enhancing predictive accuracy and model interpretability. *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*. pp. 1671-1675.

Joseph, E., Mishra, A. & Rabiu, I. 2019. Forecast on close stock market prediction using support vector machine (SVM). *International Journal of Engineering Research & Technology (IJERT)* 8(02): 37-43.

Kabir, M.R., Bhadra, D., Ridoy, M. & Milanova, M. 2025. LSTM–Transformer-based robust hybrid deep learning model for financial time series forecasting. *Sci*. 7(1): 7.

Kucher, O., Kurov, A. & Wolfe, M.H. 2023. A shot in the arm: The effect of COVID-19 vaccine news on financial and commodity markets. *Financial Review* 58(3): 575-596.

Li, J.F. & Lin, Z.C. 2024. An application of multi-factor model on anomaly in China stock market. *International Journal of Religion* 5(7): 849-856.

Li, S., Wang, Y., Zhang, Z. & Zhu, Y. 2022. Research on the factors affecting stock price volatility. *7th International Conference on Financial Innovation and Economic Development (ICFIED 2022)*. pp. 2884-2889.

Mgalla, M. 2023. The impact of the COVID-19 pandemic on the stock market returns: Evidence from the Chinese stock market. *Proceedings of the 2023 6th International Conference on Computers in Management and Business*. pp. 51-57.

Muth, C., Oravecz, Z. & Gabry, J. 2018. User-friendly Bayesian regression modeling: A tutorial with rstanarm and shinystan. *The Quantitative Methods for Psychology* 14(2): 99-119.

Nirmala, B. & Pundareeka Vittala, K.R. 2023. A study on interaction between global exchanges and Indian capital market. *International Journal of Advanced Research in Science, Communication and Technology* 3(8): 819-823.

Oanh, T.T.K. 2022. The impact of COVID-19 vaccination on stock market: Is there any difference between developed and developing countries? *Heliyon* 8(9): e10718.

Rawat, V., Singh, K. & Sahu, Y. 2023. Reliance stock market analysis with RBF kernel using support vector machine. *2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*. pp. 939-943.

Razali, N.S. & Nur-Firyal, R. 2021. Predicting index price based on the COVID-19 cases and deaths. *Sains Malaysiana* 50(9): 2833-2846.

Serkan, U., Çömlekçi, I. & Ali, Ö. 2022. Stock market reaction to COVID-19 vaccination rate: International study. *Konuralp Medical Journal* 14(S1): 183-191.

Singh, R. 2025. The role of social media sentiment in predicting stock returns: An econometric and text analytics approach. *International Journal for Multidisciplinary Research* 7(1): IJFMR250135498.

Smola, A.J. & Schölkopf, B. 2004. A tutorial on support vector regression. *Statistics and Computing* 14(3): 199-222.

Tabash, M.I., Babar, Z., Sheikh, U.A., Khan, A.A. & Anagreh, S. 2022. The linkage between oil price, stock market indices, and exchange rate before, during, and after COVID-19: Empirical insights of Pakistan. *Cogent Economics & Finance* 10(1): 2129366.

Tan, X., Ma, S., Wang, X., Zhao, Y., Wang, Z. & Xiang, L. 2022. The dynamic impact of COVID-19 pandemic on stock returns: A TVP-VAR-SV estimation for G7 countries. *Frontiers in Public Health* 10: 859647.

Wong, H.T. 2022. The impact of real exchange rates on real stock prices. *Journal of Economics, Finance and Administrative Science* 27(54): 262-276.

Yang, S. 2025. Pandemic, policy, and markets: Insights and learning from COVID-19's impact on global stock behavior. *Empirical Economics* 68: 555-583.

Yunianto, I., Haryanto, T., Mutoffar, M.M., Fadhillah, Y. & Putria, N.E. 2023. Comparison of stock price predictions using support vector regression and recurrent neural network methods. *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*. pp. 927-932.

Zhang, C. 2023. The impact of COVID-19 on the international financial market. *Highlights in Business, Economics and Management* 19: 547-551.

Zhang, D., Hu, M. & Ji, Q. 2020. Financial markets under the global pandemic of COVID-19. *Finance Research Letters* 36: 101528.

Zhao, X. 2021. Application of Bayesian regression model in financial stock market forecasting. *2nd International Conference on Management, Economy and Law (ICMEL 2021)*. pp. 140-144.

Zouaghia, Z., Kodia, Z. & Ben Said, L. 2024. Predicting the stock market prices using a machine learning-based framework during crisis periods. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-024-20270-3

*Corresponding author; email: nurfiryal@ukm.edu.my

APPENDIX A. TIME SERIES LINEAR REGRESSION RESULTS

TABLE S1. Time Series Linear Regression results for KLSE pre-vaccination

Variable	β	SE	p-value
Lag 2	0.045	0.018	0.041*
RDP5	0.062	0.027	0.019*
BSE 30	0.080	0.031	0.012*
Nikkei 225	0.053	0.024	0.029*
Intercept	-0.032	0.051	0.524

TABLE S2. Time Series Linear Regression results for KLSE post-vaccination

Variable	β	SE	p-value
Lag 1	0.038	0.014	0.035*
Lag 3	0.054	0.023	0.025*
RDP5	0.072	0.029	0.018*
Intercept	-0.025	0.043	0.510

TABLE S3. Time Series Linear Regression results for STI pre-vaccination

Variable	β	SE	p-value
Lag 1	0.050	0.016	0.038*
Lag 2	0.067	0.025	0.022*
RDP5	0.075	0.028	0.015*
BSE 30	0.045	0.021	0.030*
Nikkei 225	0.049	0.020	0.027*
Intercept	-0.030	0.048	0.515

TABLE S4: Time Series Linear Regression results for STI post-vaccination

Variable	β	SE	p-value
Lag 1	0.042	0.015	0.032*
Lag 2	0.058	0.022	0.024*
Lag 3	0.069	0.026	0.017*
Lag 4	0.041	0.018	0.029*
RDP5	0.048	0.021	0.033*
Nikkei 225	0.052	0.024	0.028*
Intercept	-0.025	0.042	0.508

TABLE S5. Time Series Linear Regression results for JKSE pre-vaccination

Variable	β	SE	p-value
Lag 2	0.048	0.017	0.037*
Lag 6	0.063	0.026	0.021*
RDP5	0.077	0.030	0.014*
BSE 30	0.046	0.023	0.027*
Intercept	-0.031	0.049	0.508

TABLE S6. Time Series Linear Regression results for JKSE post-vaccination

Variable	β	SE	p-value
Lag 2	0.039	0.014	0.034*
IDR	0.056	0.021	0.023*
Nikkei 225	0.071	0.028	0.016*
RDP5	0.044	0.019	0.028*
Intercept	-0.024	0.041	0.506

BAYESIAN REGRESSION RESULTS

TABLE S7. Bayesian Regression results for KLSE pre-vaccination

Variable	Mean	SD	95% CI
Lag 2	0.046	0.019	[0.010, 0.082]
RDP5	0.061	0.025	[0.012, 0.103]
BSE 30	0.079	0.030	[0.020, 0.130]
Nikkei 225	0.052	0.023	[0.005, 0.089]
Intercept	-0.031	0.050	[-0.102, 0.048]

TABLE S8. Bayesian Regression results for KLSE post-vaccination

Variable	Mean	SD	95% CI
Lag 1	0.037	0.013	[0.008, 0.068]
Lag 3	0.055	0.024	[0.009, 0.098]
RDP5	0.073	0.027	[0.015, 0.120]
Intercept	-0.026	0.042	[-0.095, 0.043]

TABLE S9. Bayesian Regression results for STI pre-vaccination

Variable	Mean	SD	95% CI
Lag 1	0.048	0.018	[0.012, 0.082]
Lag 2	0.066	0.026	[0.013, 0.105]
RDP5	0.074	0.029	[0.019, 0.112]
BSE 30	0.044	0.022	[0.009, 0.075]
Nikkei 225	0.051	0.021	[0.010, 0.085]
Intercept	-0.029	0.047	[-0.095, 0.046]

TABLE S10. Bayesian Regression results for STI post-vaccination

Variable	Mean	SD	95% CI
Lag 1	0.041	0.014	[0.011, 0.069]
Lag 2	0.057	0.023	[0.015, 0.098]
Lag 3	0.068	0.027	[0.018, 0.112]
Lag 4	0.040	0.019	[0.010, 0.072]
RDP5	0.047	0.022	[0.014, 0.085]
Nikkei 225	0.051	0.025	[0.009, 0.092]
Intercept	-0.024	0.041	[-0.080, 0.045]

TABLE S11. Bayesian Regression results for JKSE pre-vaccination

Variable	Mean	SD	95% CI
Lag 2	0.048	0.017	[0.013, 0.080]
Lag 6	0.064	0.027	[0.016, 0.108]
RDP5	0.078	0.031	[0.021, 0.126]
BSE 30	0.047	0.024	[0.010, 0.088]
Intercept	-0.030	0.047	[-0.091, 0.048]

TABLE S12. Bayesian Regression results for JKSE post-vaccination

Variable	Mean	SD	95% CI
Lag 2	0.041	0.016	[0.012, 0.072]
IDR	0.057	0.022	[0.014, 0.095]
Nikkei 225	0.072	0.029	[0.018, 0.121]
RDP5	0.045	0.020	[0.011, 0.081]
Intercept	-0.023	0.040	[-0.080, 0.045]

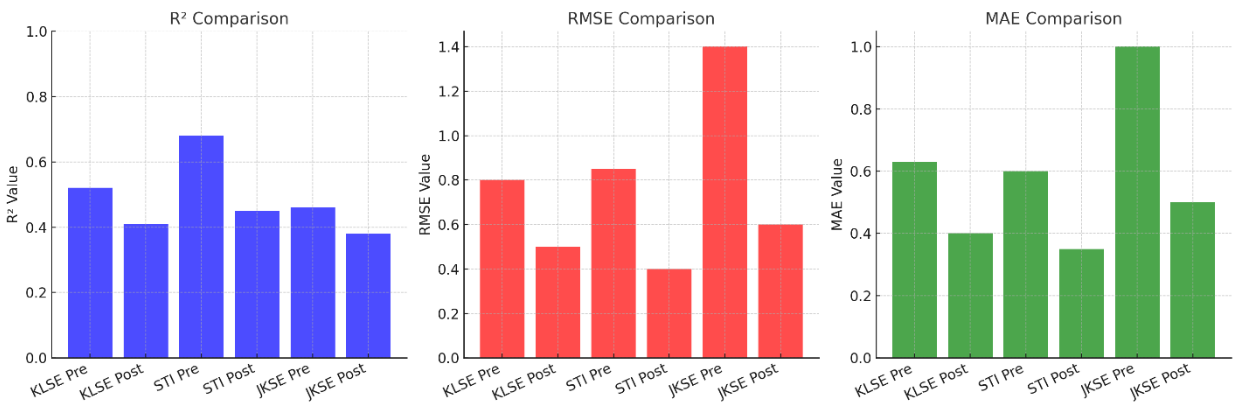


FIGURE S1. R^2, RMSE and MAE comparison across models