Sciencestatistics
Journal of Statistics, Probability, and Its Application
Available at https://scholar.ummetro.ac.id/index.php/sciencestatistics/index

Vol. 1, No. 2, 2023, 74-83
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

Bayesian Structural Time Series Model for Forecasting the Composite Stock Price Index in Indonesia

Indah Suciati[1*], Mustofa Usman[2]

[1*,2] Department of Mathematics, Universitas Lampungs, Lampung, Indonesia
*corresponding author: indahsuciati222@gmail.com

Received June 28, 2023; Received in revised form July 5, 2023; Accepted July 15, 2023

Abstrak. *Salah satu model yang dapat digunakan untuk meramalkan data deret waktu adalah model Bayesian Structural Time Series (BSTS). Model BSTS merupakan model yang lebih modern dan dapat mengatasi ketidakpastian data secara lebih baik. Dalam model BSTS, digunakan algoritma pengambilan sampel Markov Chain Monte Carlo (MCMC) untuk mensimulasikan distribusi posterior, yang menghaluskan hasil peramalan atas sejumlah besar model yang potensial menggunakan rata-rata model Bayesian. Tujuan penelitian ini adalah memperoleh model BSTS terbaik untuk data IHSG di Indonesia berdasarkan komponen states dan jumlah iterasi MCMC, serta memperoleh hasil peramalan untuk nilai IHSG di Indonesia 24 bulan ke depan yaitu periode Juli 2023 sampai dengan Juni 2024. Hasil yang diperoleh yaitu berdasarkan perbandingan nilai R-square pada model, model BSTS dengan komponen state tren linear lokal dan musiman, serta jumlah iterasi MCMC* $n = 500$ *merupakan model BSTS terbaik yang dapat digunakan untuk peramalan nilai IHSG di Indonesia dengan nilai R-square sebesar 99,96%. Adapun hasil peramalan nilai IHSG di Indonesia periode Juli 2023 sampai dengan Juni 2024 berikisar pada nilai 6589-6760, dengan nilai peramalan terendah terletak pada bulan Oktober 2023 dan nilai peramalan tertinggi terletak pada bulan Maret 2023.*

Kata kunci*: MCMC; Model Bayesian Structural Time Series; Peramalan*

Abstract. One of the models that can be used to predict time series data is the Bayesian Structural Time Series (BSTS) model. The BSTS model is a more modern model and can handle data movement better. In the BSTS model, the Markov Chain Monte Carlo (MCMC) sampling algorithm is used to simulate the posterior distribution, which smoothes the forecasting results over a large number of potential models using Bayesian averaging models. The purpose of this study was to obtain the best BSTS model for Composite Stock Price Index (CSPI) data in Indonesia based on the state component and the number of MCMC iterations, and obtain forecasting results for CSPI value in Indonesia for the next 24 months, namely the period July 2023 to June 2024. The results obtained are based on a comparison of the R-square values in the model, the BSTS model with local linear trend and seasonal state components, and the number of MCMC iterations n = 5 00 is the best BSTS model that can be used for forecasting the CSPI value in Indonesia with an R-square value of 99.96%. The results of forecasting the CSPI value in Indonesia for the period July 2023 to June 2024 range from 6589 to 6760, with the lowest forecasting value in October 2023 and the highest in March 2023.

Keywords: MCMC, Bayesian Structural Time Series Models, Forecasting

Vol. 1, No. 2, 2023, 74-83
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

INTRODUCTION

Currently, the most developed method is the forecasting method. One method that is often used to carry out forecasting is time series analysis. Time series analysis is the analysis of a set of data in a past time period which is useful for knowing or predicting future conditions (Rohmaningsih, et al., 2016). In its development, time series analysis is often used in the fields of economics and finance, especially in the capital market, namely shares.

In practice, to attract investors to invest in the capital market, good capital market conditions are needed. The indicator that is often used by investors in conducting capital market analysis before investing is the Composite Stock Price Index (IHSG) (Syahadati, et al., 2021). According to Nurwani (2016), IHSG is the combined value of company shares listed on the Indonesia Stock Exchange (BEI), whose movements indicate the current economic conditions in the capital market. According to Syahadati (2021), IHSG is one of the economic indicators used to see how the economy is in Indonesia, so it is necessary to make forecasts to help make future decisions.

One model that can be used to predict time series data is the Bayesian Structural Time Series (BSTS) model. The BSTS model can be used for forecasting, looking for related variables (feature selection), inferring causal relationships, and knowing aspects that have an impact at the moment (nowcasting) (Scott & Varian, 2014). The BSTS model is a stochastic state space model that can investigate trend, seasonal and regression components separately (Feroze, 2020). The BSTS model is a more modern model and can handle data uncertainty better. The uncertainty in the data is caused by stochastic or random movements over time so that for more accurate forecasting, a model is needed that can handle this uncertainty well. In the BSTS model, because analytical calculations of the Bayesian posterior distribution are very difficult, numerical calculations using the Markov Chain Monte Carlo (MCMC) method will be used to simulate the posterior distribution, which smooths the forecasting results over a large number of potential models using the Bayesian averaging model (George & McCulloch, 1997; Hoeting, et al., 1999; Madigan & Raftery, 1994). In this research, the bsts package in R software will be used to assist in the calculations.

Research regarding the BSTS model includes research conducted by Almarashi & Khan (2020), regarding time series modeling using BSTS on Flying Cement share price data. Then the BSTS model was compared with the classic Autoregressive Integrated Moving Average (ARIMA) model, based on forecasting plots and Mean Absolute Percent Error (MAPE). The results of this research show that for short-term forecasting, both ARIMA and BSTS are good to use, but for long-term forecasting, the BSTS model with local level components is the best model to use. Apart from that, research was conducted by Tang & Halmkrona (2022), regarding the comparison of ARIMA, BSTS, and Generalized Additive Models (GAM) models to develop a package delay forecasting model using tracking data. The results of this research show that by using the RMSE, MAE, and MASE assessment criteria, the BSTS model is able to provide better performance compared to the ARIMA and GAM models.

METHOD

Research Data

In this research, the BSTS model will be formed on IHSG data in Indonesia for the period January 1995 to June 2023 obtained via the website https://finance.yahoo.com/, So the amount of research data used was 342 data. Then the best BSTS model will be obtained based on the states component and the number of MCMC iterations, as well as forecasting

Vol. 1, No. 2, 2023, 74-83
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

results for IHSG data in Indonesia for the next 24 months, namely the period July 2023 to June 2024. The calculation process will be assisted by R software.

Structural Time Series (STS)

According to Almarashi & Khan (2020), in the Structural Time Series (STS) model, the data comes from some unobserved process known as the state space and the observed data is generated from the state space with additional noise. The state space component is responsible for generating data such as trends, seasonality, cyclicity, and the effects of independent variables which will be identified separately before being used in the STS model. The general model of STS is as follows (Scott & Varian, 2012):

$$\begin{aligned} y_t &= Z_t^T \alpha_t + e_t, \quad e_t \sim N(0, H_t), \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t, \quad \eta_t \sim N(0, Q_t). \end{aligned} \tag{1}$$

Matrix $\boldsymbol{Z_t}, \boldsymbol{T_t}, \boldsymbol{R_t}, \boldsymbol{H_t},$ and $\boldsymbol{Q_t}$ initially assumed to be known and *error* e_t and η_t assumed to be serially independent and independent of each other at all time points. Initial state vector α_1 diasumsikan $\sim N(\alpha_1, P_1)$ independently from $e_1, e_2, \dots, e_n$ and $\eta_1, \eta_2, \dots, \eta_n$, where α_1 and P_1 assumed to be known in advance.

Local Level

The local level model is the simplest model in the state space model (Almarashi & Khan, 2020). This local level model assumes the trend is a random walk. So the local level model is defined as follows:

$$\begin{aligned} y_t &= \mu_t + e_t, \quad e_t \sim N(0, \sigma_e^2), \\ \mu_{t+1} &= \mu_t + u_t, \quad u_t \sim N(0, \sigma_u^2). \end{aligned} \tag{2}$$

In the local level model equation, the structural parameters Z_t,T_t,R_t have a scalar value of 1, as well H_t in the form of constant variance σ_e^2 and Q_t in the form of constant variance σ_u^2 at the local level model. The parameter of the model is the variance of the error (σ_e^2, σ_u^2). The prior of this component depends on the parameters σ_u^2.

Local Linear Trends

The local linear trend model assumes that the mean and slope of the trend follow a random walk. According to Durbin & Koopman (2012), The local linear trend model equation is as follows (Brodersen, *et al.*, 2015):

$$\begin{aligned} y_t &= \mu_t + e_t, \quad e_t \sim N(0, \sigma_e^2), \\ \mu_{t+1} &= \mu_t + \delta_t + u_t, \quad u_t \sim N(0, \sigma_u^2) \\ \delta_{t+1} &= \delta_t + v_t, \quad v_t \sim N(0, \sigma_v^2), \end{aligned} \tag{3}$$

with:

μ_t = trend value at time to -t
δ_t = expectations of value increases μ between times t until $t + 1$.

Semi Local Linear Trends

The semi-local linear trend model is a generalization of the local linear trend model but is more useful for long-term forecasting (Almarashi & Khan, 2020). This model assumes that the level or mean component moves according to a random walk, while the slope component moves based on the AR(1) process which is centered on the non-zero potential value of D. The observation equation containing a semi-local linear trend component is as follows:

Sciencestatistics
Journal of Statistics, Probability, and Its Application
Available at https://scholar.ummetro.ac.id/index.php/sciencestatistics/index

Vol. 1, No. 2, 2023, 74-83
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

$$\begin{aligned} y_t &= \mu_t + e_t, & e_t &\sim N(0,\sigma_e^2), \\ \mu_{t+1} &= \mu_t + \delta_t + u_t, & u_t &\sim N(0,\sigma_u^2), \\ \delta_{t+1} &= D + \rho(\delta_t - D) + v_t, & v_t &\sim N(0,\sigma_v^2), \end{aligned} \quad (4)$$

with:

μ_t = trend value at time to -t
δ_t = expectations of value increases μ between times t until $t+1$
$|\rho| < 1$ = *learning rate* where local trends are updated.

Seasonal

The seasonal model can be considered as a regression with dummy variables as many as S seasons where the number of coefficients must be 1 and the expected value of the coefficient is 0 for 1 full cycle of S season (Scott & Varian, 2013). The STS model observation equation containing a seasonal component is as follows:

$$\begin{aligned} y_t &= \tau_t + e_t, \quad e_t \sim N(0,\sigma_e^2). \\ \tau_{t+1} &= -\sum_{s=0}^{S-2} \tau_{t-s} + w_t, \quad w_t \sim N(0,\sigma_w^2), \end{aligned} \quad (5)$$

with:

S = many seasons
τ_t = seasonal coefficient of joint contribution to the response variable y_t.

Seasonal effects of τ_t can be changed depending on the seasonality of the data. For example, if you have daily data, then use it $S = 7$, for quarterly data, used $S = 4$, and for monthly data used $S = 12$. Then for the weekly annual cycle effect it is used $S = 52$, because there are 52 weeks in 1 year (Durbin & Koopman, 2012).

***Bayesian Structural Time Series* (BSTS)**

The BSTS model is a model that can be used for forecasting, looking for related variables (feature selection), inferring causal relationships, and knowing aspects that have an impact at the moment (nowcasting) (Scott & Varian, 2014). The BSTS model is an STS model with a Bayesian approach in estimating the model, where the model is represented in the form of a state space model (Almarashi & Khan, 2020). The steps for applying the Bayesian approach to the STS model are:

1. Determine the prior distribution for each parameter in the model.
2. Obtain the posterior distribution. However, because the analytical calculation or integral solution of the Bayesian posterior distribution formula is very difficult. Then numerical calculations are carried out using MCMC simulation methods such as Gibbs sampling, namely by taking samples from the posterior distribution so that parameter estimation values from the BSTS model can be obtained (George & McCulloch, 1997). Where the computation is done using the bsts package in R software.

RESULT AND DISCUSSION

In this research, IHSG data was used in Indonesia for the period January 1995 to June 2023. The IHSG data used in the research is presented in Figure 1 as follows.

Vol. 1, No. 2, 2023, 74-83
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

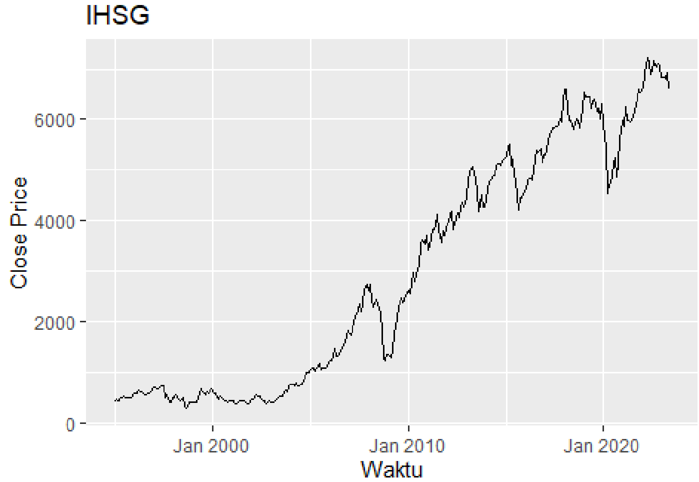

Figure 1. IHSG data in Indonesia.

Based on Figure 1 above, it can be seen that the IHSG data pattern in Indonesia is not stationary because the data fluctuates significantly and shows an up and down trend over time. There are 3 extreme points of the most significant decline in the IHSG value, namely in 2008, 2013 and 2020. This is because Indonesia was experiencing an economic crisis in that year which was caused by a recession or economic downturn due to economic activity, drastic technological developments. , as well as the COVID-19 pandemic that occurred in 2020.

Time Series Data Decomposition

Time series data can be decomposed into 4 main components, namely trend, seasonal, cyclical, and irregular or random components. The decomposition of IHSG data in Indonesia is presented in Figure 2 as follows.

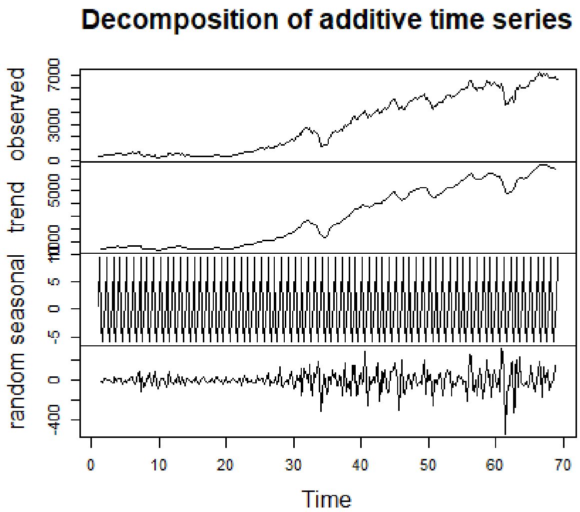

Figure 2. Decomposition of IHSG Data.

Based on Figure 2, it can be seen that the observation data pattern can be broken down into trend, seasonal and random component patterns of IHSG data in Indonesia. The trend component pattern in the data appears to rise and fall erratically. Meanwhile, the seasonal

Vol. 1, No. 2, 2023, 74-83
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

component pattern is additive because the amplitude is always constant. Most of the random component patterns in this data are almost the same, but at some times there is a significant increase in random values.

BSTS Model

Based on the results of time series decomposition on previous IHSG data, it can be seen that the data has trend and seasonal patterns, so each BSTS model is formed which contains 1 state component, including local level, local linear trend, semi-local linear trend, and contains 2 state components. namely local and seasonal levels, local and seasonal linear trends, and semi-local and seasonal linear trends. Where the number of seasons used is S=12, because the data used is monthly and annual data. Apart from that, 3 MCMC iteration values will be used, namely n=200, 500, and 1000. The performance of each BSTS model formed will be measured based on the R-square value. So that the best model can be obtained that can be used for forecasting IHSG values in Indonesia. The comparison of R-square values for each model is presented in Table 1 as follows.

Table 1. Comparison of R-square values

States component	Iteration MCMC	*R-square*	residual.sd	prediction.sd
Local Level	n=200	0.9994436	54.87818	155.0524
	n=500	0.9994570	54.21390	154.9722
	n=1000	0.9994407	55.02152	155.0751
Local Linear Trends	n=200	0.9993640	58.67236	160.4927
	n=500	0.9994632	53.90059	157.6876
	n=1000	0.9994129	56.37257	157.2835
Semi Local Linear Trends	n=200	0.9993174	60.78302	151.0270
	n=500	0.9992924	61.88421	150.9046
	n=1000	0.9992383	64.20711	150.9441
Local and Seasonal Levels	n=200	0.9994797	53.06515	158.3109
	n=500	0.9994796	53.07153	155.9893
	n=1000	0.9994763	53.23818	156.3127
Local and Seasonal Linear Trends	n=200	0.9993103	61.09698	161.6011
	n=500	**0.9996125**	45.79840	157.8433
	n=1000	0.9994655	53.78350	157.8487
Semi-Local and Seasonal Linear Trends	n=200	0.9992999	61.55837	151.6635
	n=500	0.9993784	58.00447	150.6253
	n=1000	0.9993549	59.09098	151.0108

Based on Table 1, it is known that the R-square value for each model has values that are not much different, namely ranging from 99.92% to 99.96%. Apart from that, it is known that the model with the largest R-square value, namely 99.96%, is located in the BSTS model which consists of local and seasonal linear trend state components, with the number of seasons used, namely S=12 and MCMC iterations, namely n=500. Based on the R-square value, it can be said that the model is good for use in forecasting IHSG values in Indonesia. The posterior distribution of the best BSTS model is as follows.

Sciencestatistics
Journal of Statistics, Probability, and Its Application
Available at https://scholar.ummetro.ac.id/index.php/sciencestatistics/index

Vol. 1, No. 2, 2023, 74-83
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

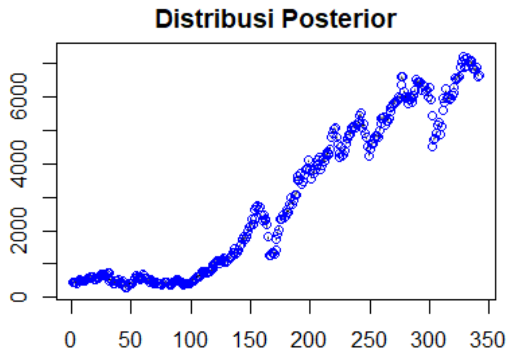

Figure 3. Decomposition of IHSG Data.

Based on Figure 3, the blue dots show the variable data used for modeling. The state components for the best BSTS model are as follows.

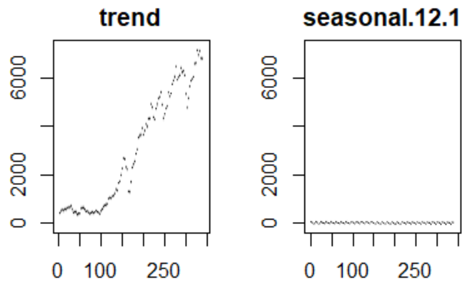

Figure 4. Components of the BSTS State Model.

Forecasting Results

Based on the best BSTS model obtained, the forecast value of IHSG data in Indonesia for the next 24 months, namely the period July 2023 to June 2024 is presented in Figure 7 as follows.

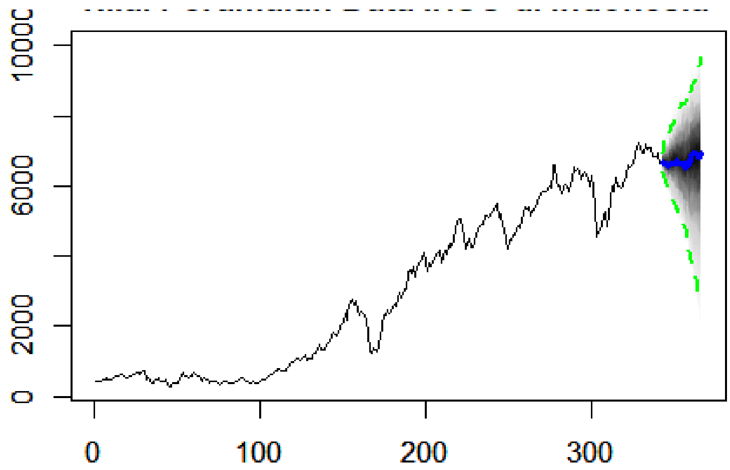

Figure 5. IHSG Value Forecast Plot.

In Figure 5 above, the dark blue line is the result of forecasting the IHSG value in Indonesia for the next 24 months. It can be seen that the forecasting value for the IHSG from July 2023

Sciencestatistics
Journal of Statistics, Probability, and Its Application
Available at https://scholar.ummetro.ac.id/index.php/sciencestatistics/index

Vol. 1, No. 2, 2023, 74-83
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

to June 2024 is quite volatile. Then the green dotted line is the IHSG forecast value which shows that after June 2023, the IHSG value will fluctuate with the IHSG value range as shown by the green line interval. This means that the IHSG value may increase or decrease in the following months. For greater clarity, a plot for forecasting values for the next 24 months is presented in Figure 8 as follows.

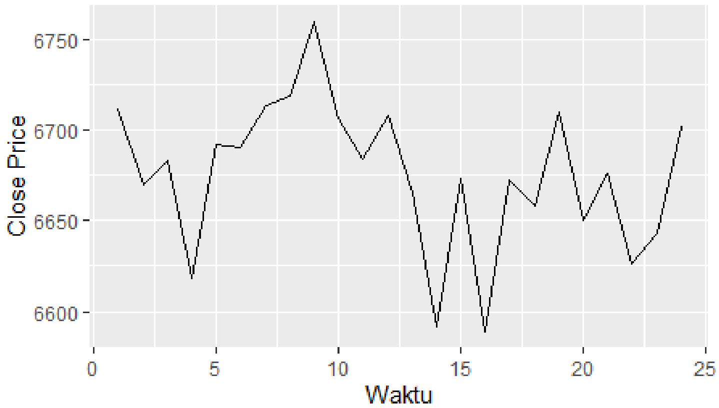

Figure 6. JCI Data Forecasting Results

Based on Figure 8 above, it is known that the forecast value of IHSG data for the next 24 months ranges from 6589-6760, with an average value of 6676. The lowest forecast value is in the 16th month, namely October 2023 and the highest forecast value is located in the 9th month, namely March 2023. For more details, a table of IHSG forecasting values in Indonesia for the next 24 along with intervals of forecasting values is presented in Table 3 as follows.

Table 3. Summary of Forecasting Results

Time	Forecasting IHSG Value	Forecasting Value Intervals
July 2023	6712	6413 - 6915
August 2023	6670	6258 - 7273
September 2023	6684	6077 - 7281
October 2023	6619	5983 - 7212
November 2023	6692	5892 - 7456
December 2023	6690	5686 - 7744
January 2024	6713	5576 - 7796
February2024	6719	5507 - 7791
March 2024	6760	5361 - 7984
April 2024	6707	5285 - 8124
May 2024	6684	5193 - 8224
June 2024	6708	5080 - 8385
July 2024	6666	5003 - 8341
August 2023	6592	4785 - 8355
September 2023	6673	4539 - 8451
October 2023	6589	4348 - 8731
November 2023	6673	3983 - 8784
December 2023	6658	3930 - 8672
January 2024	6710	3818 - 8893
February 2024	6650	3580 - 8958

Vol. 1, No. 2, 2023, 74-83
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

March 2024	6676	3342 - 9008
April 2024	6627	3005 - 9142
May 2024	6644	3045 - 9265
June 2024	6702	2746 - 9684

CONCLUSION

The best BSTS model for forecasting IHSG in Indonesia is the BSTS model which consists of local and seasonal linear trend state components, with the number of seasons used, namely S=12 and MCMC iterations, namely n=500. Then, we obtained the results of forecasting the IHSG value in Indonesia for the next 24 months, namely the period July 2023 to June 2024, ranging between values 6589-6760, with the lowest IHSG forecast value located in the 16th month, namely October 2023 and the highest IHSG forecast value located in the month the 9th, namely March 2023.

REFERENCE

Almarashi, A.M. & Khan, K. (2020). Bayesian Structural Time Series. *Nanoscience and Nanotechnology Letters*. 12: 54-61.

Brodersen, K.H., Gallusser, F., Koehler, J., Remy, N., & Scott, S.L. (2015). Inferring Causal Impact using Bayesian Structural Time-Series Models, *The Annals of Applied Statistics*. 9(1): 247–274.

Durbin, J. & Koopman, S.J. (2012). *Time Series Analysis by State Space Methods*. 2nd Edition. Oxford University Press, UK.

Feroze, N. (2020). Forecasting the patterns of COVID-19 and Causal Impacts of Lockdown in Top Five Affected Countries Using Bayesian Structural Time Series. *Elsevier Journal of Chaos, Solitons, and Fractals*. 140: 1-8.

George, E.I. & McCulloch, R.E. (1997). Approaches for Bayesian Variable Selection. *Statistica Sinica*. 7: 339–374.

Ghahramani, S. (2016). *Fundamentals of Probability with Stochastic Processes*. 3rd Edition. CRC Press, US.

Hazhiah, T.I., Sugito, & Rahmawati, R. (2012). Estimasi Parameter Distribusi Weibull Dua Parameter Menggunakan Metode Bayes. *Media Statistik*. 5(1): 27-35.

Hoeting, J.A., Madigan, D., Raftery, A.E., & Volinsky, C.T. 1999. Bayesian Model Averaging: A Tutorial. *Statistical Science*. 14(4): 382-417.

Hutahaean, H.D. (2018). Analisa Simulasi Monte Carlo untuk Memprediksi Tingkat Kehadiran Mahasiswa dalam Perkuliahan. *Jurnal of Informatic Pelita Nusantara*. 3(1): 41-45.

Koerniawan, V., Nilsen, A., Sari, F.P., Ayyasy, M.Y., & Indranto, S.P. (2022). Pemodelan Peluang Transisi Rantai Markov dengan Simulasi Monte Carlo Berdasarkan Distribusi Multinoulli untuk Memprediksi Harga Indeks Saham. *Jurnal Statistika dan Aplikasinya*. 6(2): 276-287.

Madigan, D. & Raftery, A.E. (1994). Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window. *Journal of the American Statistical Association*. 89(428): 1535–1546.

Mokilane, P., Debba, P., Yadavalli, V.S.S., & Sigauke, C. (2019). Bayesian Structural Time Series Approach to a Long-Term Electricity Demand Forecasting. *Applied Mathematics & Information Sciences*. 13(2): 189-199.

Vol. 1, No. 2, 2023, 74-83
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

Nurwani. (2016). Analisis Pengaruh Inflasi, Nilai Tukar Rupiah, dan Suku Bunga SBI Terhadap Pergerakan Indeks Harga Saham Gabungan di Bursa Efek Indonesia. *Jurnal Ilmu Ekonomi* dan *Studi Pembangunan*. 16(2).

Rohmaningsih, N.M., Sudarno, & Diah Safitri. (2016). Pemodelan dan Peramalan Volatilitas pada Return Saham Bank Bukopin Menggunakan Model Asymmetric Power Autoregressive Conditional Heteroskedasticity (APARCH). *Jurnal Gaussian*. 5(4): 705-715.

Ross, S.M. (2010). *Introduction to Probabilit Models*. 10th Edition. Elsevier, California.

Scott, S.L. & Varian, H.R. (2014). Predicting the Present with Bayesian Structural Time Series. *Int. J. Mathematical Modelling and Numerical Optimisation*. 5(1-2): 4-23.

Simatupang, S. (2022). Simulasi Monte Carlo dalam Memprediksi Ketersediaan Barang PT. Terang Abadi Pekanbaru. *Jurnal Sistem Informasi* dan *Manajemen*. 10(1): 176-184.

Syahadati, A., Lengkong, N.C., Safitri, O., Machsus, S., Putra, Y.R., & Nooraeni, R. (2021). Analisis Sentimen Penerapan PSBB di DKI Jakarta dan Dampaknya Terhadap Pergerakan IHSG. *Jurnal Teknoinfo*. 15(1): 20-25.

Tang, A. & Halmkrona, J. (2022). Applying Software Engineering Principles to Develop Parcel Delay Forecasting Models using Tracking Data: A Study of the models ARIMA, BSTS, and GAM. Tesis. Department of Computer Science and Engineering, Chalmers University of Technology, Sweden.