

COMPARISON OF SETAR AND LSTAR MODELS IN PREDICTING REGULAR CONVENTIONAL GASOLINE PRICE DATA

Anastasya A. M. Sidabalok¹, Subian Saidi², Bernadhita Herindri Samodera Utami^{3*}, Netti Herawati⁴

^{1,2,3,4}Department of Mathematics, Universitas Lampung, Bandar Lampung, Lampung, Indonesia

*Corresponding Author; Email: bernadhita@fmipa.unila.ac.id

ABSTRACT

Forecasting involves predicting what may happen in the future. An essential part of this process is examining time series data, especially when it shows nonlinear characteristics. This study concentrates on predicting the prices of standard regular gasoline in the United States, which consists of a time series affected by numerous economic and seasonal elements. It utilizes two nonlinear time series models: Self-Exciting Threshold Autoregressive (SETAR) and Logistic Smooth Transition Autoregressive (LSTAR). SETAR segments the data into distinct regimes based on a threshold variable, whereas LSTAR allows more gradual changes between regimes through a logistic function. Past research has indicated that both models effectively capture structural shifts in financial and economic time series. By assessing the forecasting performance of the SETAR and LSTAR models, this research seeks to identify which model better forecasts gasoline price fluctuations that demonstrate nonlinear behavior. The findings are anticipated to enhance understanding of energy market trends and aid in making better economic choices.

Keywords: Time Series, Forecasting, SETAR, LSTAR, Gasoline Prices, Nonlinear Models

1. Introduction

Forecasting is a crucial tool for supporting effective and efficient planning, especially in the fields of economics and business, where decision-making plays a critical role (Brockwell & Davis, 2002). Generally, forecasting is categorized into three-time horizons: short-term, medium-term, and long-term. Short-term forecasting refers to predictions made over several days to a few months; medium-term forecasting typically covers one to two years, and long-term forecasting projects trends over several years into the future (Granger & Jeon, 2007; Krisanti et al., 2024).

One commonly applied approach in forecasting is time series analysis, which involves examining data recorded at consistent time intervals to identify patterns and generate future projections (Wei, 2006). A relevant example of time series forecasting is the analysis of conventional regular gasoline prices in the United States. According to data from the Federal Reserve Economic Data (FRED), this type of gasoline, with an octane rating of around 87, is widely consumed due to its affordability. However, its price is subject to fluctuations caused by various factors such as global crude oil prices, distribution costs, taxation, and seasonal demand. These fluctuations exhibit

nonlinear patterns over time, making accurate forecasting essential for guiding economic strategies and shaping effective energy policies.

To handle nonlinear patterns, researchers often use the SETAR (Self-Exciting Threshold Autoregressive) and LSTAR (Logistic Smooth Transition Autoregressive) models. SETAR divides data into distinct regimes based on threshold values. The process of determining the SETAR model involves the number of regimes (m), autoregressive order (p), delay parameters, and threshold variables. The j -regime SETAR (d, p_1, \dots, p_j) model is defined as follows (Tongan & Booij, 2023).

$$Z_t = \begin{cases} \phi_{0,1} + \sum_{n=1}^{p_1} \phi_{n,1} Z_{t-n} + \varepsilon_t, & Z_{t-d} \leq r_1 \\ \phi_{0,2} + \sum_{n=1}^{p_2} \phi_{n,2} Z_{t-n} + \varepsilon_t, & r_1 < Z_{t-d} \leq r_2 \\ \vdots & \\ \phi_{0,j} + \sum_{n=1}^{p_j} \phi_{n,j} Z_{t-n} + \varepsilon_t, & r_{j-1} < Z_{t-d} \end{cases}$$

LSTAR uses a logistic function to model smoother transitions between regimes. The LSTAR model in simple form can be written as follows:

$$X_t = \phi'_1 Y_t \left(1 - \left(\frac{1}{1 + \exp(-\gamma(S_t - c))} \right) \right) + \phi'_2 Y_t \left(\frac{1}{1 + \exp(-\gamma(S_t - c))} \right) + \varepsilon_t$$

where $S_t = X_{t-l}$ with l is the delay parameter l which is a positive integer with $l > 0$. The parameter c is known as the threshold while γ indicates the speed and smoothness of the transition (Terasvirta & Granger, 1993).

Previous studies have shown the effectiveness of these models. For instance, Krisanti *et al.* reported that the LSTAR model achieved a low AIC value, indicating a good model fit (Krisanti *et al.*, 2024). Similarly, Tongan and Booij found that the SETAR model outperformed the K-Nearest Neighbors (K-NN) method by yielding a lower RMSE (Tongan & Booij, 2016). Building on these findings, the present study aims to compare the performance of the SETAR and LSTAR models in forecasting conventional gasoline prices. By analyzing their ability to capture nonlinear patterns and regime shifts, this research seeks to determine which model produces more accurate forecasting results.

2. METHODS AND MATERIAL

The data used in this study are secondary time series data taken from the Federal Reserve Economic Data (FRED). This data source can be accessed through the link <https://fred.stlouisfed.org/series/GASREGCOVW>. This dataset contains weekly regular conventional gasoline prices for the period 16 February 2015 to 10 February 2025, with a total of 533 observations. The steps taken are as follows:

- 1) Plotting regular conventional gasoline price data,
- 2) Looking at data stationarity related to variance using Box-Cox transformation, as well as to the average through the Augmented Dickey-Fuller test,
- 3) Conduct nonlinearity test,
- 4) Identifying the SETAR model,
- 5) Identify the LSTAR model
- 6) Comparing the accuracy of the SETAR model with the LSTAR model seen from the smallest AIC value and MAPE value,
- 7) Perform forecasting using the selected model on regular conventional gasoline price data for the next 10 periods by comparing it with 10 validation data.

3. RESULTS AND DISCUSSION

Plot Identification of Observation Data

The results of the regular conventional gasoline price data plot are as follows:

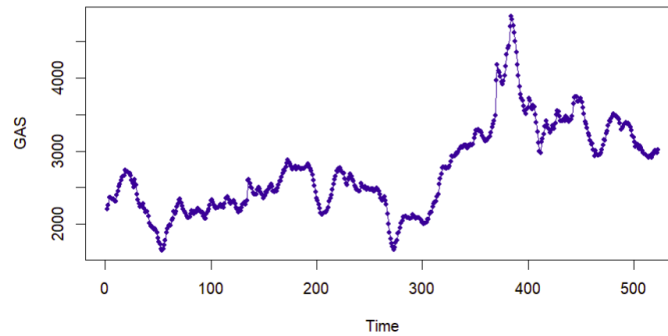


Figure 1. Regular Conventional Gasoline Price Data

Figure 1 shows that the price of regular conventional gasoline in the United States fluctuates significantly during the observation period. The highest increase occurs around time point 400, possibly influenced by external factors such as energy policy or global oil prices. After that, the price drops sharply and exhibits high volatility. Overall, there is no consistent long-term trend, indicating that gasoline prices tend to be influenced by short-term, external changes.

Data Stationarity Check

In this study, the Box-Cox transformation test and the Augmented Dickey-Fuller (ADF) test are conducted to evaluate whether the regular conventional gasoline price data is stationary in terms of its variance and average. The results of the Box-Cox transformation of the regular traditional price of gasoline data are as follows:

Table 1. Box-Cox Transformation

(λ)	1
---------------	---

The Box-Cox lambda (λ) value of 1 indicates that the data meets the requirements for variance stationarity. However, to ensure overall stationarity, an ADF test on the average regular conventional gasoline price data is necessary.

Table 2. Augmented Dickey-Fuller (ADF) Test

Augmented Dickey-Fuller Test (t -statistics)	p -value
-2.7277	0.2703

Based on Table 2, the p -value (0.2703) $> \alpha$, so accept H_0 . This indicates that the regular conventional gasoline price data is not stationary. Therefore, a differencing process is required on the average. Below are the ADF test results after differencing once:

Table 3. One-Time Differencing Augmented Dickey-Fuller (ADF) Test

Augmented Dickey-Fuller Test (t -statistics)	p -Value
-7.0856	0.01

Based on Table 3, the p -value (0.01) $< \alpha$, so reject H_0 . This indicates that the regular conventional gasoline price data is stationary.

Nonlinearity Test

Nonlinearity tests can be done using the Terasvirta test. This test aims to identify whether the regular conventional gasoline price data contains nonlinearity or not, as follows:

Table 4. Nonlinearity Test Results

F	p -value
17.779	3.406e-08

Based on Table 6, the nonlinearity test results show that the p -value ($3.406e - 08$) $< \alpha$ (0.05). Thus, it can be concluded that rejecting H_0 which means that regular conventional gasoline price data is nonlinear. Before entering into SETAR and LSTAR modeling, data division will be carried out, namely 523 data used as training data and 10 data used as validation data.

SETAR Modeling

The SETAR model is identified by determining the autoregressive order (p), number of regimes (j), delay length (d), and threshold value (r). This process also involves embedding dimension (m) and time distance (τ), which are determined based on the minimum entropy value. The following table presents the embedding dimension and entropy results.

Table 5. Embedding Dimension Value and Time Distance

Time Distance	Embedding Dimension	Entropy
1	3	0.9891891
1	4	0.9786215
1	5	0.9552358
1	6	0.9741173
1	7	0.9946277

Based on the table, the optimal embedding dimension is 5, with the lowest entropy of 0.9552358 and a time distance of 1. Next, the number of regimes is determined through a nonlinearity test against the threshold by comparing the p -value against the significance level of 0.05. The test results are presented in the following table:

Table 6. Nonlinearity vs Threshold Test Results

Test	F Test Statistics	p -value
Linear AR vs 1 Threshold SETAR	46.99551	0.001
Linear AR vs 2 Threshold SETAR	66.44195	0.001
1 Threshold SETAR vs 2 Threshold SETAR	17.82317	0.300

Table 6 shows that the data fit the one threshold SETAR (2-regime) model, because the linear AR vs. one threshold and linear AR vs. two threshold tests produce a p -value < 0.05 , while the one threshold vs. two threshold test produces a p -value > 0.05 . Therefore, the best model is SETAR which has one threshold. Furthermore, the

parameters of delay length (d), the threshold value (r), and AR order for each regime (p_1, p_2) are determined based on the smallest AIC value, using embedding dimension five and delay one.

Table 7. AIC Value of 2-Regime SETAR Model

Delay	p_1	p_2	r	AIC
0	1	3	27	4019.412
0	1	4	27	4019.553
0	1	4	28	4019.612
0	1	3	28	4019.765
0	1	4	26	4020.306
0	2	3	27	4021.333
0	2	4	27	4021.473
0	1	3	48	4021.477
0	1	5	27	4021.518
0	2	4	28	4021.523

Based on Table 7, the best model is the 2-regime SETAR (0,1,3), with a threshold of 27 and the lowest AIC of 4019.412. This model is used for parameter estimation and significance testing, the results of which are presented in the next table.

Table 8. SETAR Model Parameter Estimation Results

Coefficient	Estimation	p -value	AIC	MAPE
$\phi_{0,1}$	6.482699	0.0254908	4017	13.43 %
$\phi_{1,1}$	0.777328	$< 2.2\text{e-}16$		
$\phi_{0,2}$	16.922545	0.0116557		
$\phi_{1,2}$	0.276478	0.0005112		
$\phi_{2,2}$	-0.205751	0.0064303		
$\phi_{3,2}$	0.369436	$7.7\text{e-}05$		

Based on Table 8, it can be seen that the SETAR (0,1,3) model has a p -value of each significant coefficient (p -value < 0.05). Then, the equation for the 2-regime SETAR (0,1,3) model is as follows:

$$Z_t = \begin{cases} 6.482699 + 0.777328Z_{t-1} + \varepsilon_t, & Z_t \leq 27 \\ 16.922545 + 0.276478Z_{t-1} - 0.205751Z_{t-2} + 0.369436Z_{t-3} + \varepsilon_t, & Z_t > 27 \end{cases}$$

LSTAR Modeling

The next stage is LSTAR modeling, which begins with Box-Jenkins model identification through ACF and PACF plots to determine the ARMA model, followed by parameter estimation, diagnostic tests, and LSTAR parameter estimation.

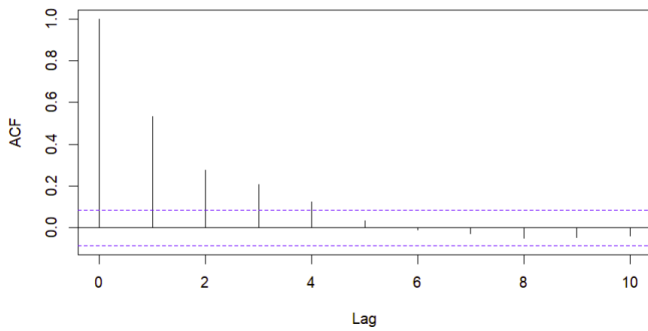


Figure 2. ACF Plot of Regular Conventional Gasoline Price Data

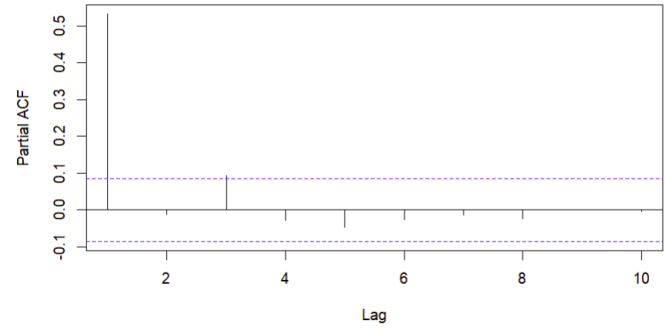


Figure 3. PACF Plot of Regular Conventional Gasoline Price Data

From the analysis of the ACF and PACF plots shown to form the conjecture model in Table 9, the next step is to estimate the parameters of the model. This process can be done by looking at the probability values of the parameters that have been determined or by finding the smallest AIC value.

Table 9. Box-Jenkins Model

No	Box-Jenkins Model	AIC
1	ARMA(1,0)	5539.04
2	ARMA(0,1)	5566.35
3	ARMA(1,1)	5540.94
4	ARMA(2,0)	5540.98
5	ARMA(2,1)	5541.40

Based on Table 9, the ARMA(1,0) model is chosen as the best model because it has the lowest AIC value. Once the model is determined, diagnostic tests are conducted to ensure that the residuals fulfill the basic assumptions of the model. These tests include the Ljung-Box test for autocorrelation, the Kolmogorov-Smirnov test for normality, and the ARCH test to detect heteroscedasticity. The results of these tests are important to assess the feasibility of the model before it is used in LSTAR modeling.

Table 10. Ljung-Box Test Results

Model	X-squared	p-value
ARMA(1,0)	110.09	0.3226

Based on Table 10, the Ljung-Box test results show that the ARMA(1,0) model has a p -value (0.3226) $> \alpha$ (0.05). Therefore, it can be concluded that it does not reject H_0 , which means there is no autocorrelation between residuals.

Table 11. Normality Test Results

Model	D _{count}	p-value
ARMA(1,0)	0.0747981	0.05878

The normality test results, based on Table 11, show that the ARMA (1,0) model has a p -value (0.05878) $> \alpha$ (0.05). So, it can be concluded that do not reject H_0 which means that the residuals in the data are normally distributed.

Table 12. Heteroscedasticity Test Results

Model	Chi-Squared	P-Value
ARMA(1,0)	16.446	0.1716

The heteroscedasticity test results, based on Table 12, show that the ARMA(1,0) model has a p -value (0.1716) $> \alpha$ (0.05). Therefore, it can be concluded that we do not reject H_0 which means there is no heteroscedasticity.

The next step is to estimate the parameters of the LSTAR model by determining the model order (m), which is determined based on the results of the Partial Autocorrelation Function (PACF) plot analysis in Figure 3. Based on the results of the study, the relevant order is at the 1st and 3rd lags. Therefore, parameter estimation for the LSTAR model is performed using orders $m = 1$ and $m = 3$, assuming a delay length of 1 and two transitions. The parameter estimates of the LSTAR model are presented in the following table.

Table 13. Parameter Estimation of LSTAR Model

Model	Parameter	Estimation	p -value	AIC	MAPE
LSTAR(1,1)	$\phi_{0,1}$	5.062159	0.02782	4037	11.72 %
	$\phi_{1,1}$	0.745199	$<2.2\text{e-}16$		
	$\phi_{2,0}$	8.001653	0.48236		
	$\phi_{2,1}$	-0.481552	$5.901\text{e-}06$		
	γ	20.013021	0.95644		
	c	59.491659	$8.123\text{e-}11$		
LSTAR(3,1)	$\phi_{0,1}$	6.4555902	0.0249661	4026	12.75 %
	$\phi_{1,1}$	0.7876944	$<2\text{e-}16$		
	$\phi_{1,2}$	-0.0178908	0.7743062		
	$\phi_{1,3}$	-0.0020537	0.9659175		
	$\phi_{2,0}$	10.4670380	0.1487183		
	$\phi_{2,1}$	-0.5112176	$2.839\text{e-}06$		
	$\phi_{2,2}$	-0.1878595	0.0538198		
	$\phi_{2,3}$	0.3714895	0.0003544		
	γ	20.0044117	0.9539811		
	c	27.5367700	0.0015540		

From Table 13, the results of model estimation and evaluation show that LSTAR(1,1) is selected as the best model. Although the LSTAR(3,1) model has a lower AIC value, the LSTAR(1,1) model provides more accurate prediction results (MAPE 11.72%), and most of its parameters are statistically significant. Therefore, the equation of the LSTAR(1,1) model can be expressed as follows:

$$X_t = (5.062159 + 0.745199X_{t-1}) \left(1 - \left(\frac{1}{1 + \exp(20.013021(X_{t-1} - 59.491659))} \right) \right) + (8.001653 - 0.481552X_{t-1}) \left(\frac{1}{1 + \exp(20.013021(X_{t-1} - 59.491659))} \right) + \varepsilon_t$$

Model Comparison

From the analysis results between SETAR and LSTAR models, the next step is to choose the best model by referring to the smallest AIC value and the minimum MAPE value. The comparison of SETAR and LSTAR models is shown as follows:

Table 14. Comparison of SETAR and LSTAR Models

Model	AIC	MAPE
SETAR(0,1,3)	4017	13.43 %
LSTAR(1,1)	4026	11.72 %

Based on Table 14, the LSTAR(1,1) model has a lower MAPE value (11.72%) compared to the SETAR(0,1,3) model (13.43%), which indicates that the LSTAR(1,1) model provides more accurate forecasting results. Although the AIC value of the SETAR(0,1,3) model is smaller (4017), in the context of forecasting, a higher level of accuracy is preferred. Therefore, the LSTAR(1,1) model is selected as the best model to predict the price of regular conventional gasoline.

Forecasting

The best model chosen in this study is the LSTAR(1,1) model based on forecasting accuracy criteria, namely by looking at the smallest MAPE value. Next, the price of regular conventional gasoline is forecasted for the next 10 periods, from the 524th to the 533rd period. These forecasting results are then validated by comparing them against the last 10 actual data from the dataset, which is used as validation data. The comparison between the actual data and the forecasting results is shown in Table 15.

Table 15. Forecasting Results of Regular Conventional Gasoline Price Data 10 Periods ahead

Period	Actual Data	Forecasting Results	MAPE
524	3.021	3.064	1.43%
525	2.997	3.101	3.46%
526	2.951	3.133	6.17%
527	2.947	3.162	7.30%
528	2.943	3.189	8.36%
529	2.999	3.214	7.17%
530	3.040	3.238	6.51%
531	3.118	3.261	4.57%
532	3.043	3.284	7.90%
533	3.020	3.304	9.40%
Average MAPE			6.23%

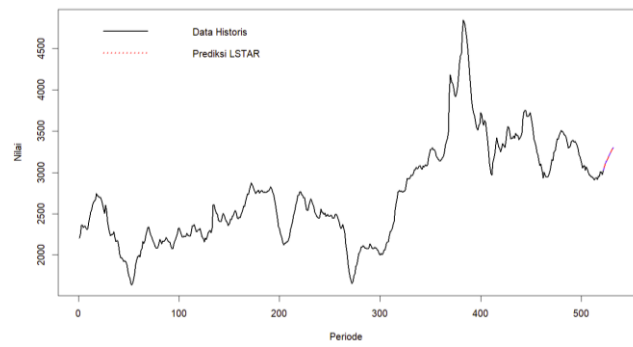


Figure 4. Results of Data Forecast of Regular Conventional Gasoline Price 10 Periods ahead

Based on Table 15 and Figure 4, the forecasting results in 10 validation periods show that the predicted value is quite close to the actual data, such as in the 524th period with actual data of \$3,021/gal and forecasting results of \$3,064/gal, and the 525th period with actual data of \$2,997/gal and forecasting results of \$3,101/gal. This shows that the LSTAR(1,1) model has a good ability to follow the movement pattern of regular conventional gasoline prices, making it suitable for forecasting purposes.

4. CONCLUSION

Based on the analysis results, the SETAR(0,1,3) and LSTAR(1,1) models are equally capable of representing nonlinear patterns in regular conventional gasoline price data. The SETAR model provides a lower AIC value of 4017, which indicates a statistically good model fit. However, in the context of forecasting, accuracy is a key consideration. The LSTAR(1,1) model has the advantage of a lower MAPE value of 11.72% compared to SETAR, which recorded a MAPE of 13.43%. Therefore, despite the higher AIC value, the LSTAR(1,1) model is considered more appropriate for forecasting purposes.

In addition, the prediction results for the 10 validation periods show that the LSTAR(1,1) model produces forecasting values that are close to the actual data. For example, in the 524th and 525th periods, the difference between the actual and predicted values is very small, indicating that the model is quite responsive to price movement patterns. This proves that LSTAR(1,1) is not only able to capture nonlinear dynamics but also provides good prediction performance, making it reliable in supporting decision-making regarding future fuel prices.

REFERENCES

- [1] C. T. Krisanti, N. Herawati, A. Sutrisno, N. Nusyirwan. 2024. Logistic Smooth Transition Autoregressive (LSTAR) and Exponential Smooth Transition Autoregressive (ESTAR) Methods in Predicting the Exchange Rate of Farmers in Lampung Province, Indonesia. *International Journal of Multidisciplinary Research and Analysis*, 7(07): 3209–3215. DOI: 10.47191/ijmra/v7-i07-18
- [2] C. W. J. Granger, and Y. Jeon. 2007. Long-term Forecasting and Evaluation. *International Journal of Forecasting*, 23(4):539-551. DOI: 10.1016/j.ijforecast.2007.07.002
- [3] D. K. Shangodoyin, O. A. Adebile, and R. Arnab. 2006. Spesification of Logistic Smooth Transition Autoregressive Model: An Alternatif Representation of the Transition Function. *Journal of Mathematics and Statistic*. 5(4):251-256. DOI: 0.3844/jmssp.2009.251.259
- [4] D. N. Gujarati and D. C. Porter. 2009. *Basic Econometrics*. 5th ed. New York, USA: McGraw Hill Irwin.

- [5] Federal Reserve Bank of St. Louis. 2015. *Conventional gasoline prices: West Coast (PADD 5)*. [Online]. Available: <https://fred.stlouisfed.org/series/GASREGCOVW>. [Accessed Jan. 19, 2025].
- [6] G. E. P. Box and G. L. Jenkins. 1976. *Time Series Analysis Forecasting and Control*. 1st ed. San Francisco, USA: Holden Day.
- [7] H. Tongal, and M. J. Booi. 2016. A Comparison of Nonlinear Stochastic Self-Exciting Threshold Autoregressive and Chaotic k-Nearest Neighbour Models in Daily Streamflow Forecasting. *Water Resour Manage*. 30: 1515–1531. DOI: 10.1007/s11269-016-1237-6
- [8] I. Azzam. 2007. The effect of model-selection uncertainty on autoregressive models estimates. *International Research Journal of Finance and Economics*. 2 (11): 79–92. (Available at SSRN: <https://ssrn.com/abstract=1908225>)
- [9] P. J. Brockwell, and R. A. Davis. 2002. *Introduction to Time Series and Forecasting*. 2th ed. Springer, New York.
- [10] R. S. Tsay. 2010. *Analysis of Financial Time Series*. 3rd ed. Canada: John Wiley & Sons.
- [11] T. Teräsvirta. 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*. 89 (425): 208–218. DOI: 10.1080/01621459.1994.10476462
- [12] T. Teräsvirta, C. F. Lin, and C. W. J. Granger. 1993. Power of the neural networks linearity test. *Journal of Time Series Analysis*. 14 (2): 209–220. DOI: 10.1111/j.1467-9892.1993.tb00139.x
- [13] V. Dijk. 1999. *Smooth Transition Models: Extensions and Outliers Robust Inference*. Amsterdam, Netherlands: Tinbergen Institute.
- [14] W. S. Chan, A. C. S. Wong, and H. Tong. 2004. Some nonlinear threshold autoregressive time series models for actuarial use. *North American Actuarial Journal*. 8 (4): 37–61. DOI: 10.1080/10920277.2004.10596170.
- [15] W. W. S. Wei. 2006. *Time Series Analysis: Univariate and Multivariate Methods*. USA: Addison-Wesley Publishing Company.

Cite this Article

Anastasya A. M. Sidabalok, Subian Saidi, Bernadhita Herindri Samodera Utami, Netti Herawati, “**Comparison Of Setar And Lstar Models In Predicting Regular Conventional Gasoline Price Data**”, *International Journal of Multidisciplinary Research in Arts, Science and Technology (IJMRAST)*, ISSN: 2584-0231, Volume 3, Issue 6, pp. 01-10, June 2025.

Journal URL: <https://ijmrast.com/>

DOI: <https://doi.org/10.61778/ijmrast.v3i6.139>



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).