Received: 21 Jan 2026 | Revised: 02 Mar 2026 | Accepted: 25 Mar 2026 | Available online: 31 May 2026
https://doi.org/10.13057/ijas.v9i1.114966

E-ISSN: 2621-086X
https://jurnal.uns.ac.id/ijas

Indonesian Journal of Applied Statistics Vol. 9, No.1, pp. 1-15, 2026

Forecasting the U.S. Treasury Yield Curve Using the Hybrid Dynamic Nelson-Siegel and Long Short-Term Memory (LSTM) Method

Neni Alya Firdausanti*, **Alya Shafira**
Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
*Corresponding Author. E-mail: neni.alya@its.ac.id

Abstract

U.S. Treasury (UST) securities are widely regarded as safe-haven assets and serve as global financial benchmarks, making the U.S. Treasury yield curve a key indicator of market expectations and economic risks, including recession probabilities. For Indonesia, where foreign exchange reserves are partly allocated to UST securities, accurate yield curve forecasts are essential for effective reserve management and monetary policy formulation. This study proposes a hybrid forecasting framework that integrates the dynamic Nelson–Siegel (DNS) model with long short-term memory (LSTM) networks to improve the accuracy and stability of U.S. Treasury yield curve forecasts. The decay parameter in the DNS model is estimated using the Newton–Raphson method, while the remaining parameters are estimated using ordinary least squares (OLS). The resulting DNS latent factors are subsequently used as input features for the LSTM model under various hyperparameter configurations. Forecasting performance is evaluated using the root mean squared error (RMSE) and benchmarked against a DNS–ARIMA model. The empirical results demonstrate that the proposed DNS-LSTM approach consistently outperforms DNS-ARIMA across all maturities, yielding lower forecasting errors and greater flexibility in capturing yield curve dynamics, particularly during the post-pandemic period. Overall, the DNS-LSTM model offers a more robust and data-driven alternative to traditional yield curve forecasting methods. These findings have practical implications for foreign reserve management, exchange rate stabilization, and investment decision-making. Future research may extend this framework by incorporating macroeconomic variables and exploring longer forecast horizons.

Keywords: Bonds, yield, dynamic Nelson-Siegel, long short-term memory, U.S. Treasury.

How to Cite:
N. A. Firdausanti and A. Shafira, "Forecasting the U.S. Treasury yield curve using the hybrid dynamic Nelson-Siegel and long short-term memory (LSTM) method," *Indonesian Journal of Applied Statistics,* vol. 9, no. 1, pp. 1-15, 2026, doi: 10.13057/ijas.v9i1.114966.

1. INTRODUCTION

Foreign exchange reserves play a pivotal role in maintaining a nation's macroeconomic stability. Beyond fulfilling international obligations, they serve as instruments for exchange rate stabilization and monetary policy implementation, thereby contributing significantly to national economic resilience. In Indonesia, the management of foreign exchange reserves is mandated to Bank Indonesia, as stipulated in Law No. 4 of 2023 concerning the Development and Strengthening of the Financial Sector and Bank Indonesia Board of Governors Regulation No. 4 of 2024. These regulations emphasize the importance of professional and sustainable reserve management in supporting the stability of the national financial system.

In practice, the management of foreign exchange reserves is inseparable from the challenges posed by increasingly complex global financial market dynamics, including shifts in monetary policy in advanced economies, volatility in international capital flows, and heightened geopolitical uncertainty. One of the primary instruments in foreign reserve portfolios is U.S. Treasury securities, which are widely characterized as safe-haven assets with high liquidity [1],[2]. Consequently, fluctuations in U.S. Treasury yields constitute a crucial factor in determining the value and performance of foreign exchange reserves. The U.S. Treasury yield curve, particularly its slope component, is widely used as a leading indicator of global economic conditions, including recession probabilities [3].

To strengthen reserve management strategies grounded in timely and relevant market information, accurate forecasts of U.S. Treasury yields are essential. Various approaches have been developed in the academic literature, including the Nelson–Siegel model [4] and its extension by Diebold and Li [5] into the dynamic Nelson–Siegel (DNS) framework, which represents the yield curve using three latent factors: level, slope, and curvature [5]. The DNS–ARIMA model has been widely applied due to its flexibility in capturing the cross-sectional shape of the yield curve and has previously been used to forecast Indonesian government bond yields with reasonably accurate results [6]. To further enhance forecasting accuracy, state-space approaches have also been introduced in subsequent studies [7]. More recently, machine learning-based methods, particularly Long Short-Term Memory (LSTM) networks, have demonstrated strong performance in capturing nonlinear patterns and complex temporal dependencies in financial time series data.

Nevertheless, the integration of the dynamic Nelson–Siegel framework with LSTM networks remains relatively underexplored, particularly in the context of foreign reserve management by central banks in developing economies such as Indonesia. This study therefore aims to develop a U.S. Treasury yield curve forecasting model by integrating the dynamic Nelson–Siegel framework with LSTM neural networks. Under this approach, the DNS latent factors, which capture yield curve dynamics, are modeled nonlinearly using an LSTM architecture to improve forecasting accuracy and stability.

This study contributes to the literature in three main aspects. First, it provides a technical foundation for Bank Indonesia to manage foreign reserve portfolios in a more adaptive and responsive manner to global market dynamics. Second, it enriches the academic literature on yield curve modeling by proposing a hybrid approach that integrates econometric methods with machine learning techniques. Third, it offers strategic insights for investors and policymakers in formulating investment strategies and risk mitigation measures under volatile market conditions. Ultimately, the findings of this study support more robust and sustainable foreign reserve management and contribute to strengthening Indonesia's long-term economic resilience.

2. METHODS

2.1. Framework

The methodological framework of this study is illustrated in Figure 1, which provides an overview of the research workflow from data preparation to yield curve forecasting. The analysis begins with U.S. Treasury yield data, followed by the estimation of dynamic Nelson–Siegel (DNS) parameters to characterize the yield curve structure. These parameters are subsequently forecast using both classical time-series and machine learning approaches, namely ARIMA and LSTM, and substituted back into the DNS model to generate yield forecasts.

2.2. Data Collection

This study uses monthly U.S. Treasury yield data covering the period from February 2006 to February 2025, obtained from the official website of the U.S. Department of the Treasury. The dataset includes yields with maturities of 1, 2, 3, 5, 7, 10, 20, and 30 years. The sample is divided into in-sample (80%) and out-of-sample (20%) periods for model estimation and evaluation.

https://doi.org/10.13057/ijas.v9i1.114966
E-ISSN: 2621-086X

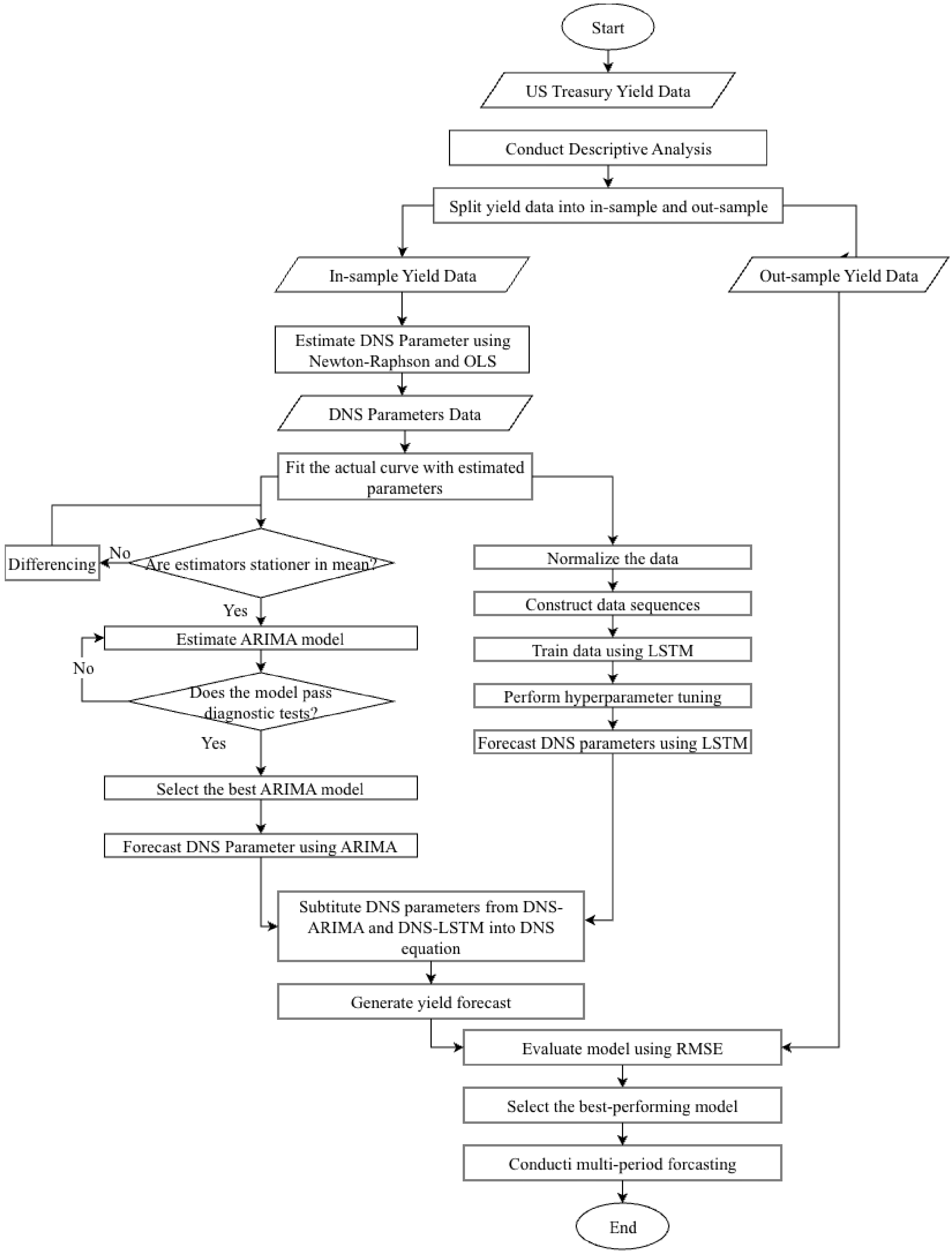

Figure 1. Research framework illustrating data preparation, dynamic Nelson–Siegel parameter estimation, and yield curve forecasting using ARIMA and LSTM models.

https://doi.org/10.13057/ijas.v9i1.114966
E-ISSN: 2621-086X

2.3. Dynamic Nelson-Siegel

The dynamic Nelson–Siegel (DNS) model represents the yield curve using three latent factors: level ($\boldsymbol{\beta_0}$), slope ($\boldsymbol{\beta_1}$), and curvature ($\boldsymbol{\beta_2}$). The decay parameter $\boldsymbol{\lambda}$ is assumed to be constant and is estimated using numerical method by minimizing the average root mean squared error (RMSE) across maturities. Given the estimated value of $\boldsymbol{\lambda}$, the time-varying DNS parameters are obtained using Ordinary Least Squares (OLS) [5].

2.4. Autoregressive Integrated Moving Average (ARIMA) Benchmark Models

To provide a benchmark for comparison, univariate ARIMA and SARIMA models are estimated for each DNS parameter following the Box–Jenkins procedure. The adequacy of each specification is evaluated based on the fundamental assumptions of ARIMA modeling [9]:

1. Stationarity in mean, assessed using the Augmented Dickey–Fuller (ADF) test, with differencing applied when necessary to eliminate unit roots.
2. Ca Correct model identification, guided by ACF and PACF patterns.
3. Statistical significance of autoregressive and moving-average parameters, evaluated at the 5% significance level.
4. Residual independence, examined using the Ljung–Box test to ensure absence of autocorrelation (white noise condition).
5. Variance stability, assessed using the ARCH-LM test to detect conditional heteroskedasticity.
6. Residual normality, evaluated using the Kolmogorov–Smirnov test.

It is important to distinguish between stationarity in mean and stationarity in variance, as they correspond to different aspects of the stochastic process. ARIMA models specify the conditional mean process, whereas ARCH-type behavior relates to the conditional variance of the residuals. Accordingly, stationarity in mean (absence of unit roots) is a prerequisite for model identification under the Box–Jenkins framework [9], while variance stability is evaluated during the diagnostic stage.

The ARCH-LM test was applied to the stationary (differenced) DNS parameters to examine conditional heteroskedasticity. Significant ARCH effects indicate time-varying volatility (volatility clustering), a common characteristic of financial time series. However, the presence of conditional heteroskedasticity affects the second moment (variance) of the process and does not invalidate the conditional mean specification of ARIMA models. While heteroskedasticity may reduce efficiency of statistical inference, it does not bias forecasts of the conditional mean

Models that failed parameter significance or residual independence were excluded from further consideration. Although residual normality and strict homoskedasticity are desirable properties, they are not imposed as strict requirements in forecasting applications, where predictive accuracy and residual independence are the primary concerns [10]. Final model selection is based on a combination of statistical adequacy, Akaike information criterion (AIC), and out-of-sample RMSE performance. These benchmark models serve solely as comparative references for evaluating the forecasting performance of the proposed DNS–LSTM framework.

2.5. Long Short-Term Memory

A long short-term memory (LSTM) neural network is employed to model nonlinear and long-term temporal dependencies in the DNS parameters [11]. The LSTM architecture is illustrated in Figure 2. The three DNS parameters are jointly forecast under a direct multi-horizon forecasting framework. Prior to model training, the data are normalized and transformed into sequential input-output structures.

https://doi.org/10.13057/ijas.v9i1.114966
E-ISSN: 2621-086X

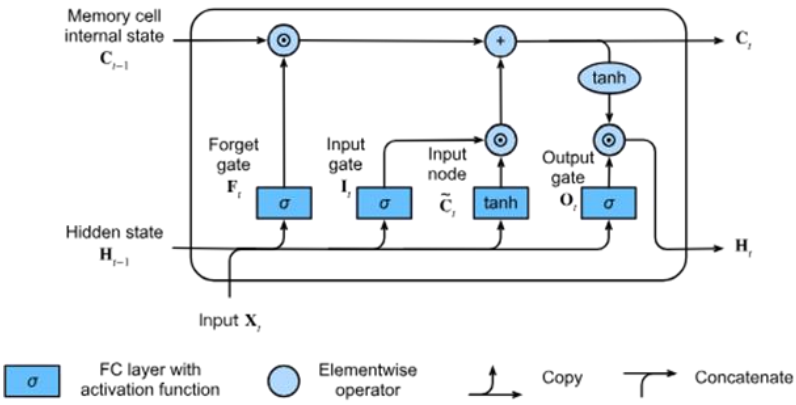

Figure 2. Architecture of the long short-term memory (LSTM) network illustrating the input, forget, and output gates used for forecasting the dynamic Nelson–Siegel parameters [12].

A grid-search strategy is conducted over key hyperparameters, including input sequence length, number of hidden units, dropout and recurrent dropout rates, learning rate, and batch size, as summarized in Table 1. Model selection is based on the lowest validation RMSE. The selected LSTM model is evaluated using a rolling (walk-forward) one-step-ahead forecasting scheme.

Table 1. Hyperparameter search space for the LSTM model tuning

Hyperparameter	Candidate Values
Input sequence length (time step)	{18, 23, 24}
Number of hidden units	{32, 64}
Dropout rate	{0.1, 0.2, 0.3}
Recurrent dropout rate	{0.0, 0.1}
Learning rate	{0.001, 0.005}
Batch size	{8, 16, 32}

In addition to the tunable hyperparameters, several training-related settings were held fixed across all model configurations to ensure comparability of results. All LSTM models were trained using the Adam optimizer with a mean squared error loss function. The maximum number of training epochs was set to 500, with early stopping based on validation loss to prevent overfitting. A plateau-based learning rate reduction scheme was applied to improve convergence stability. The LSTM layer used the default hyperbolic tangent (tanh) activation function for cell state updates and sigmoid activation for gating mechanisms. The output layer employed a linear activation function, appropriate for continuous regression forecasting. Model weights were initialized using the default Glorot uniform initializer in TensorFlow and all input variables were standardized prior to training to ensure numerical stability.

The selected LSTM model was subsequently evaluated on the test set using a rolling (walk-forward) one-step-ahead forecasting scheme. At each test period, forecasts were generated using only information available up to the preceding time point, thereby replicating a realistic real-time forecasting environment and ensuring a fair comparison with classical time-series benchmark models. Importantly, the rolling evaluation was conducted using the fixed hyperparameter configuration identified during validation, without further re-tuning or re-estimation.

2.6. Model Evaluation

A commonly used performance metric is the root mean square error (RMSE), which measures the average squared difference between actual and predicted values [13].

https://doi.org/10.13057/ijas.v9i1.114966
E-ISSN: 2621-086X

3. RESULTS AND DISCUSSION

3.1. Descriptive Statistics

This section provides an initial overview of the dataset characteristics prior to parameter estimation using the Nelson–Siegel model and subsequent forecasting with ARIMA and LSTM. The descriptive statistics summarize measures of central tendency and dispersion, as well as minimum and maximum yield values for each maturity over the observation period (Table 2).

Table 2. Descriptive statistics of U.S. Treasury yields across maturities.

Statistic	Tenor (Year)							
	$y_{t(1)}$	$y_{t(2)}$	$y_{t(3)}$	$y_{t(5)}$	$y_{t(7)}$	$y_{t(10)}$	$y_{t(20)}$	$y_{t(30)}$
Mean	0.0170	0.0180	0.0194	0.0227	0.0256	0.0282	0.0330	0.0341
Std	0.0185	0.0166	0.0152	0.0131	0.0119	0.0111	0.0108	0.0097
Min	0.0005	0.0011	0.0011	0.0021	0.0039	0.0055	0.0098	0.0120
50%	0.0066	0.0102	0.0142	0.0189	0.0228	0.0267	0.0313	0.0327
Max	0.0546	0.0516	0.0513	0.0510	0.0511	0.0515	0.0535	0.0521

The observed minimum and maximum yield values highlight periods of heightened economic stress, which are further illustrated in Figure 3.

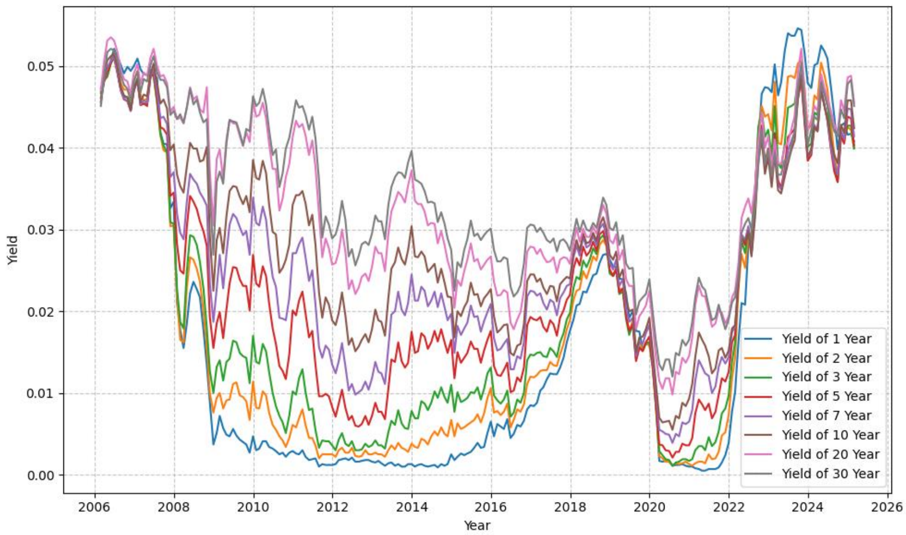

Figure 3. Historical U.S. Treasury yield curves across maturities over the sample period.

Descriptive statistics indicate that short-term yields exhibit higher volatility than medium- and long-term maturities, reflecting their greater sensitivity to monetary policy changes. Periods of pronounced yield fluctuations correspond to major economic events, including the global financial crisis, the COVID-19 pandemic, and the post-pandemic monetary tightening cycle.

3.2. Dynamic Nelson-Siegel Modeling

Parameter estimation typically assumes a constant decay parameter λ, which is obtained using the Newton–Raphson method. However, because the optimal value of λ may vary across bond categories [14], an alternative approach following Rostan [15] is adopted, whereby λ is selected by minimizing the average RMSE across maturities. Using this approach, the estimated decay parameter $\lambda = 0.59776$ corresponds to the lowest average RMSE. This value is subsequently used to estimate the time-varying parameters β_0, β_1, and β_2 using ordinary least squares (OLS), based on U.S. Treasury yield training data from February 2006 to April 2021 (Table 3).

https://doi.org/10.13057/ijas.v9i1.114966
E-ISSN: 2621-086X

Table 3. Estimated dynamic Nelson–Siegel parameters ($\hat{\beta}_{0t}, \hat{\beta}_{1t}, \hat{\beta}_{2t}$) with fixed decay parameter $\hat{\lambda} = 0.59776$.

Date	t	$\hat{\beta}_{0t}$	$\hat{\beta}_{1t}$	$\hat{\beta}_{2t}$
28-02-2006	1	0.045743	0.002694	-0.001788
31-03-2006	2	0.050034	-0.000706	-0.005654
28-04-2006	4	0.053229	-0.002218	-0.012095
31-05-2006	5	0.053279	-0.000516	-0.009957
⋮	⋮	⋮	⋮	⋮
29-01-2021	180	0.021555	-0.014588	-0.044899
26-02-2021	181	0.025474	-0.021448	-0.041717
31-03-2021	182	0.028426	-0.026100	-0.040916
30-04-2021	183	0.026975	-0.024932	-0.038568

The estimated parameters are then visualized in Figure 4.

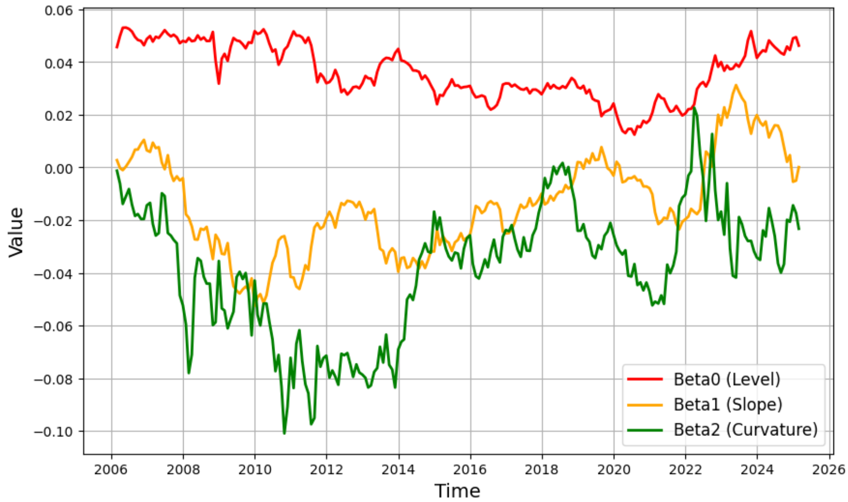

Figure 4. Time series of estimated dynamic Nelson–Siegel parameters ($\hat{\beta}_{0t}, \hat{\beta}_{1t}$, and $\hat{\beta}_{2t}$).

Figure 4 presents the estimated time series of the dynamic Nelson–Siegel estimated parameters, $\hat{\beta}_{0t}$ (level), $\hat{\beta}_{1t}$ (slope), and $\hat{\beta}_{2t}$ (curvature), for U.S. Treasury yields over the period 2006 to 2025. As shown in Figure 4, the level factor remains relatively stable over time, while the slope and curvature factors exhibit greater variability, particularly during periods of economic stress.

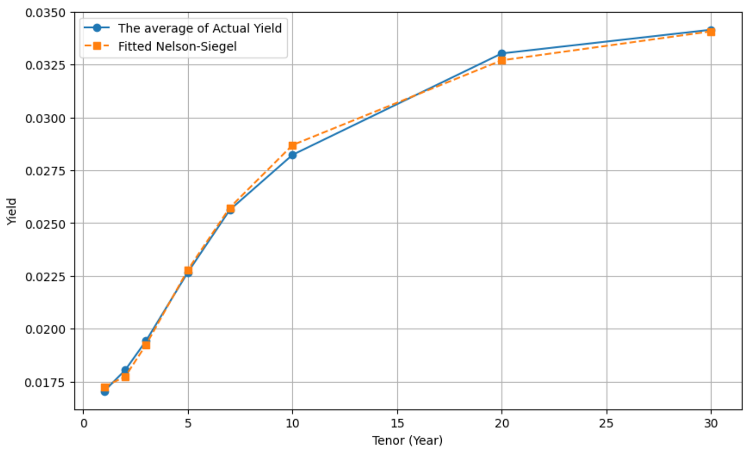

Figure 5. In-sample yield curve fitting and average U.S. Treasury yield curve based on the dynamic Nelson–Siegel model with $\lambda = 0.59776$.

https://doi.org/10.13057/ijas.v9i1.114966
E-ISSN: 2621-086X

Figure 5 illustrates the in-sample fitting performance of the dynamic Nelson-Siegel model by comparing the average observed U.S. Treasury yields across maturities with the yields implied by the estimated model. The fitted curve closely follows the shape of the average yield curve, indicating that the DNS model with the estimated decay parameter $\lambda = 0.59776$ provides an adequate representation of the term structure in the sample period. Minor deviations at certain maturities reflect approximation errors inherent in parsimonious yield curve models.

3.3. ARIMA Benchmark Model

Prior to ARIMA modeling, the stationarity of the estimated dynamic Nelson–Siegel parameters was assessed using the augmented Dickey–Fuller test. All parameters were found to be non-stationary in levels but became stationary after first differencing, confirming that the differenced series satisfy the stationarity requirement for ARIMA modeling.

In addition, as an exploratory diagnostic analysis, variance stationarity was examined using the ARCH-LM test applied to the differenced series. The purpose of this test is to detect potential conditional heteroskedasticity, which reflects time-varying volatility commonly observed in financial time series.The results indicate heterogeneous variance behavior across factors. Specifically, the ARCH-LM statistic for the differenced level factor ($\Delta\hat{\beta}_0$) is 56.6957 ($p<0.001$), indicating significant conditional heteroskedasticity. Similarly, the curvature factor ($\Delta\hat{\beta}_2$) exhibits ARCH effects (LM = 28.5951, p = 0.0045). In contrast, the slope factor ($\Delta\hat{\beta}_1$) does not display significant ARCH behavior (LM = 13.9079, p = 0.3066), suggesting variance stationarity.

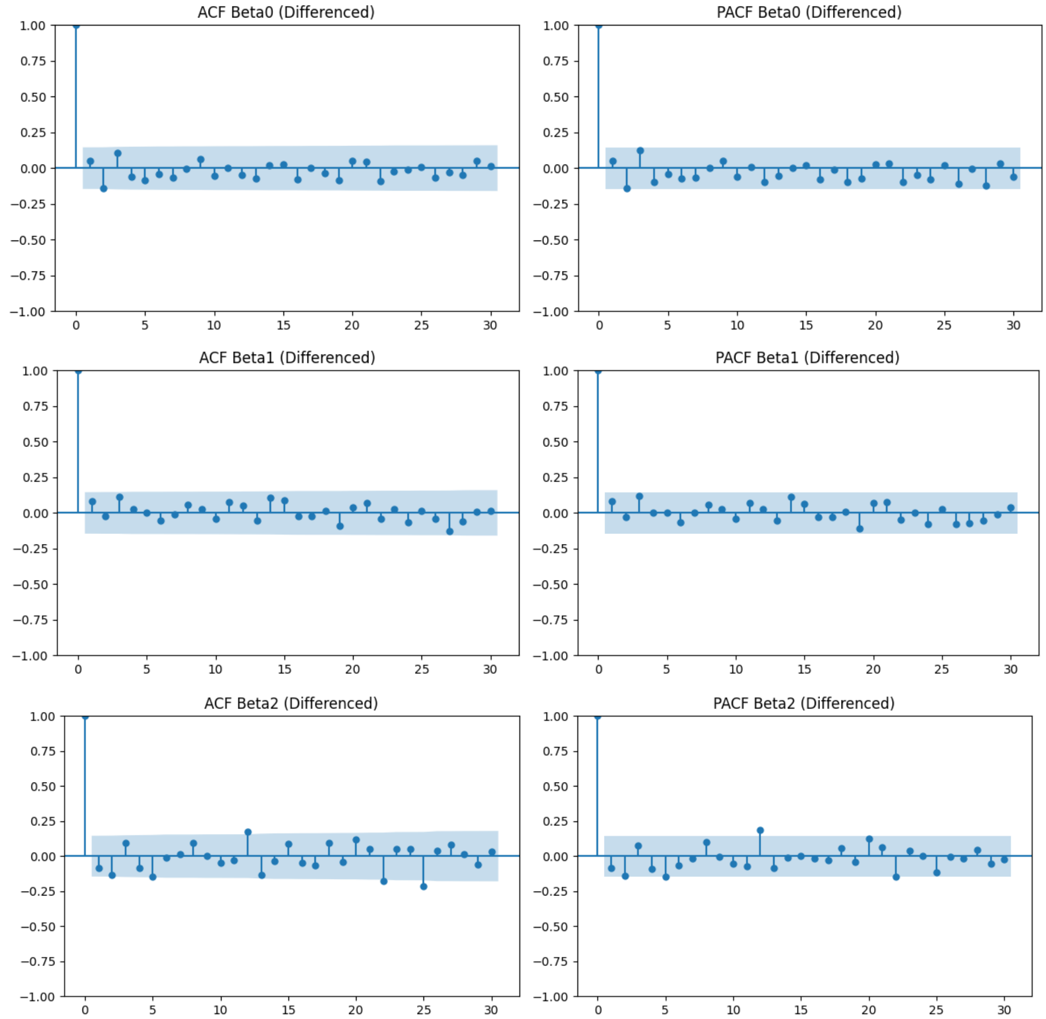

Figure 6. ACF and PACF PACF plots for estimated parameters $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$

https://doi.org/10.13057/ijas.v9i1.114966
E-ISSN: 2621-086X

It is important to note that variance stability is evaluated during the diagnostic stage rather than as a prerequisite for ARIMA model identification. Following the standard Box–Jenkins framework [9], stationarity in mean is required prior to model estimation, while tests for conditional heteroskedasticity are conducted as residual diagnostics. The presence of ARCH effects does not invalidate the conditional mean specification of ARIMA models, although it may affect inference efficiency. Since the ARIMA models in this study are used solely as benchmark forecasting tools rather than for structural inference, strict homoskedasticity is not imposed as a mandatory requirement. Consistent with Hyndman and Athanasopoulos [10], forecasting adequacy primarily depends on residual independence and predictive performance.

Examination of the autocorrelation structure of the differenced series in Figure 6 indicates that the level and slope factors exhibit no significant autocorrelation, while the curvature factor displays a seasonal pattern at lag 12. Based on these observations, appropriate ARIMA and SARIMA specifications were selected following standard Box–Jenkins criteria and employed solely as benchmark models for comparison with the proposed DNS–LSTM approach.

Table 4. Parameter significance test results for ARIMA and SARIMA models.

Parameter	Model	ARIMA Parameter	Estimate	P-Value
$\hat{\beta}_0$	ARIMA (1,1,0)	ϕ_1	0.0528	0.194
	ARIMA (0,1,1)	θ_1	0.0661	0.043
$\hat{\beta}_1$	ARIMA (1,1,0)	ϕ_1	0.0841	0.208
	ARIMA (0,1,1)	θ_1	0.0888	0.067
$\hat{\beta}_2$	ARIMA (1,1,0)	ϕ_1	-0.0814	0.216
	ARIMA (0,1,1)	θ_1	-0.0842	0.212
	SARIMA (0,1,0)(1,0,0)12	Φ_1	0.1807	0.004
	SARIMA (0,1,0)(0,0,1)12	Θ_1	0.1170	0.066
	SARIMA (0,1,0)(1,0,1)12	Φ_1	0.3215	0.449
		Θ_1	-0.1537	0.729

Since the Random Walk model has no AR or MA parameters, significance testing was not applicable. Estimation results and parameter significance are presented in Table 4. In contrast to purely predictive machine learning models, ARIMA modeling relies on statistically meaningful autoregressive and moving-average components. Therefore, only models with statistically significant parameters (p-value < 0.05) were considered for further evaluation. Models with insignificant parameters were excluded to ensure model interpretability and statistical adequacy. In addition to ARIMA and SARIMA specifications, a random walk (RW) model was also considered for further evaluation. The Random Walk model assumes that future values are equal to the most recent observation, implying no predictable structure beyond persistence.

Residual diagnostic checks were conducted to assess the adequacy of the ARIMA and SARIMA benchmark models. The results indicate that the selected benchmark specifications satisfy the white noise assumption, while some non-seasonal ARIMA models for the curvature factor ($\hat{\beta}_2$) exhibited residual autocorrelation and were therefore excluded from further analysis. Although minor deviations from normality were observed for the level factor ($\hat{\beta}_0$), residual normality is not a strict requirement in forecasting applications, where predictive accuracy is the primary objective rather than statistical inference, provided that residuals are uncorrelated and homoscedastic [10]. Given that these models are employed solely as benchmarks for comparison with the proposed DNS–LSTM framework, omitting detailed diagnostic tables does not affect the validity of the comparative results.

https://doi.org/10.13057/ijas.v9i1.114966
E-ISSN: 2621-086X

Table 5. Model selection results for ARIMA and SARIMA xpecifications based on AIC and out-of-sample RMSE for the estimated DNS parameters ($\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$).

Parameter	Model	AIC	RMSE
$\hat{\beta}_0$	ARIMA (0,1,0)	-1655.475	0.0141
	ARIMA (1,1,0)	-1653.988	0.0141
	ARIMA (0,1,1)	-1654.191	0.0142
$\hat{\beta}_1$	ARIMA (0,1,0)	-1533.417	0.0321
	ARIMA (1,1,0)	-1532.706	0.0320
	ARIMA (0,1,1)	-1532.799	0.0320
$\hat{\beta}_2$	ARIMA (0,1,1)	-1281.048	0.0246
	SARIMA (0,1,0)(1,0,0)12	-1285.156	0.0263
	SARIMA (0,1,0)(0,0,1)12	-1284.375	0.0255
	SARIMA (0,1,0)(1,0,1)12	-1283.330	0.0264

3.4. ARIMA Model Comparison

Models that failed to satisfy basic residual diagnostics were excluded from consideration. In particular, the ARIMA(0,1,0) and ARIMA(1,1,0) specifications for $\hat{\beta}_2$ were rejected due to their failure to pass the white noise test, indicating the presence of residual autocorrelation.

Based on Table 5, the selected models were ARIMA(0,1,0) for $\hat{\beta}_0$ and $\hat{\beta}_1$. For $\hat{\beta}_2$, SARIMA(0,1,0)(1,0,0)12 was chosen as the best model due to its lowest AIC, despite minor RMSE differences. Using the selected model, the out-of-sample testing for this model is presented in Figure 7.

Figure 7. Out-of-sample forecasts of the dynamic Nelson–Siegel latent factors ($\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$) generated by the selected ARIMA and SARIMA benchmark models.

Figure 7 shows out-of-sample forecasts generated by the ARIMA benchmark models exhibit overly smooth parameter trajectories and fail to capture short-term dynamics in the DNS parameters, highlighting the limitations of linear time-series models in representing nonlinear yield curve behavior.

3.5. LSTM

The dynamic Nelson–Siegel (DNS) parameters β_0, β_1, and $\beta_2$were forecast using a long short-term memory (LSTM) network, which was selected for its ability to model nonlinear dynamics and long-term

https://doi.org/10.13057/ijas.v9i1.114966
E-ISSN: 2621-086X

temporal dependencies in multivariate time series. The LSTM was implemented under a direct multi-horizon forecasting framework, allowing the joint prediction of future DNS parameter trajectories. Model hyperparameters were selected exclusively based on validation performance using a chronological training–validation split, ensuring that the test data remained strictly out of sample.

The results in Table 6 indicate that LSTM configurations with longer input windows and moderate regularization tend to achieve superior validation performance. In particular, the best-performing model employs a time step of 24 and moderate dropout regularization, suggesting that longer historical contexts combined with controlled regularization are effective in capturing the medium- to long-term dynamics of the yield curve factors.

Table 6. Top 5 LSTM hyperparameter configurations ranked by validation RMSE under the direct multi-horizon forecasting framework

Rank	Time Step	Units	Dropout	Recurrent Dropout	Learning Rate	Batch Size	Validation RMSE
1	24	64	0.2	0.1	0.005	8	0.008844
2	12	64	0.2	0.1	0.005	16	0.009138
3	12	32	0.1	0.1	0.005	16	0.009318
4	24	32	0.1	0.1	0.005	8	0.009461
5	24	64	0.1	0.0	0.001	16	0.009488

The selected LSTM model was subsequently evaluated on the test set using a rolling (walk-forward) one-step-ahead forecasting scheme. At each test period, forecasts were generated using only information available up to the preceding time point, thereby replicating a realistic real-time forecasting environment and ensuring a fair comparison with classical time-series benchmark models. Importantly, the rolling evaluation was conducted using the fixed hyperparameter configuration identified during validation, without further re-tuning or re-estimation.

Figure 8. Actual and predicted dynamic Nelson–Siegel parameters ($\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$) across training, validation, and rolling out-of-sample test periods using the selected LSTM model.

https://doi.org/10.13057/ijas.v9i1.114966
E-ISSN: 2621-086X

Figure 8 shows that the LSTM model closely tracks the observed DNS parameters during the training period and produces stable forecasts during validation, indicating good in-sample fit and generalization. In the rolling out-of-sample test period, the model captures the broad directional movements of the level and slope factors, although short-term fluctuations are partially smoothed during periods of heightened volatility. The curvature factor remains more challenging to forecast due to its higher variability; however, the LSTM is able to reproduce its medium-term dynamics without generating erratic behavior. Overall, the results suggest that the LSTM effectively learns the common dynamics of the DNS parameters while prioritizing stability over short-term noise in out-of-sample forecasting.

Table 7. RMSE of LSTM-based forecasts for dynamic Nelson–Siegel parameters across training, validation, and rolling out-of-sample test periods.

Parameter	Train RMSE	Validation RMSE	Test RMSE (Rolling)
$\hat{\beta}_0$	0.002914	0.006736	0.008218
$\hat{\beta}_1$	0.005504	0.003215	0.025735
$\hat{\beta}_2$	0.009241	0.010511	0.013664

Table 7 reports the RMSE values of the LSTM-based forecasts for each dynamic Nelson–Siegel parameter across the training, validation, and rolling out-of-sample test periods. The results indicate low training and validation RMSE values for all parameters, suggesting effective learning of the underlying temporal structure without evidence of overfitting. As expected, forecasting errors increase in the rolling out-of-sample test period. The level factor $\hat{\beta}_0$ remains relatively stable, exhibiting only a moderate increase in RMSE, while the slope factor $\hat{\beta}_1$ shows substantially higher test errors, reflecting its greater sensitivity to abrupt shifts in monetary policy expectations and market conditions. The curvature factor $\hat{\beta}_2$ displays intermediate behavior, with variability higher than the level factor but more predictable dynamics than the slope component.

3.6. Best Model Selection

To evaluate predictive performance for yield forecasting, DNS-ARIMA and DNS-LSTM models were compared across tenors using root mean square error (RMSE). A smaller RMSE indicates better predictive accuracy. The comparison results are shown in Table 8.

Table 8. RMSE comparison between DNS–ARIMA and DNS–LSTM models across all maturities.

Tenor	RMSE	
	ARIMA	LSTM
1 Tahun	0.0401	0.0311
2 Tahun	0.0368	0.0266
3 Tahun	0.0335	0.0234
5 Tahun	0.0278	0.0191
7 Tahun	0.0243	0.0167
10 Tahun	0.0208	0.0145
20 Tahun	0.0188	0.0119
30 Tahun	0.0161	0.0110

Results in Table 8 indicate that DNS-LSTM consistently produced lower RMSE values across all maturities, demonstrating its superior predictive performance in modeling the US Treasury yield curve.

3.7. Future Forecasting

The final step was forecasting DNS-LSTM parameters for six months ahead (March–August 2025) to evaluate the model's ability to project US Treasury yield dynamics. The projected DNS parameter estimates for six months ahead are shown in Table 9, as follows.

https://doi.org/10.13057/ijas.v9i1.114966
E-ISSN: 2621-086X

Table 9. Six-month-ahead projected dynamic Nelson–Siegel parameter estimates.

Date	$\hat{\beta}_{0t}$	$\hat{\beta}_{1t}$	$\hat{\beta}_{2t}$
31-03-2025	0.039594	-0.023652	-0.028137
30-04-2025	0.040549	-0.026012	-0.024224
31-05-2025	0.039775	-0.027145	-0.025956
30-06-2025	0.039741	-0.024081	-0.022959
31-07-2025	0.039140	-0.024645	-0.018839

These parameter estimates were then used in the Nelson-Siegel equation with $\hat{\lambda}$ = 0.59776 to generate yield projections for tenors from 1 to 30 years, as shown in Figure 9. The projections indicate a gradual increase in yields, reflecting market expectations regarding monetary policy and future macroeconomic conditions.

Figure 9 shows a downward yet still upward-sloping yield curve, with long-term yields higher than short-term yields. This reflects market expectations of an improving economy, monetary easing, lower inflation, or stronger demand for Treasuries. These findings provide positive signals for investors and are relevant for Bank Indonesia in managing foreign reserves and maintaining external sector stability amid global uncertainty. The vertical dashed line indicates the forecast origin. Forecasts extend beyond the last available observed data point; therefore, no realized values are available for the forecast horizon at the time of analysis.

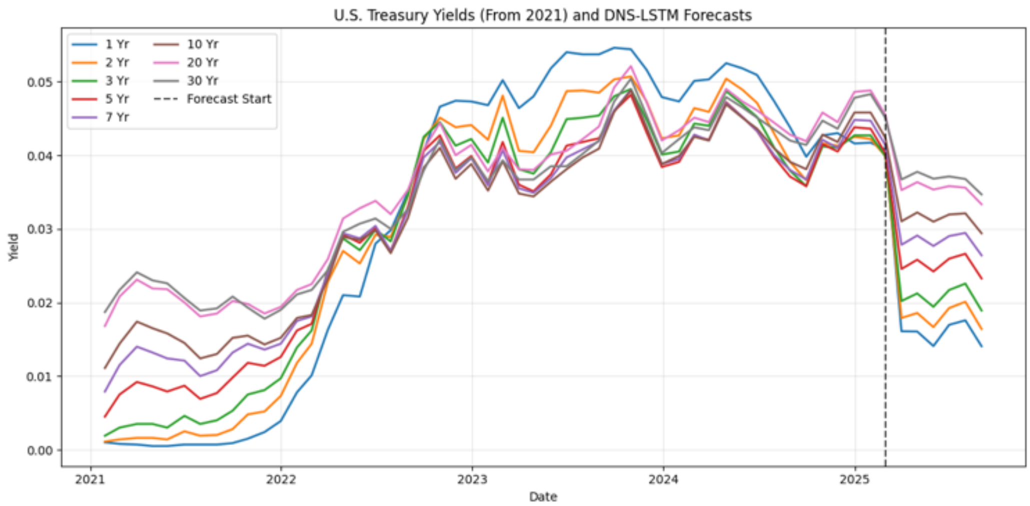

Figure 9. Projected U.S. Treasury yield curves generated using the DNS–LSTM model.

The results demonstrate that integrating the dynamic Nelson-Siegel (DNS) framework with long short-term memory (LSTM) networks improves the forecasting performance of the U.S. Treasury yield curve by capturing both its cross-sectional structure and temporal dynamics. While the DNS model provides a parsimonious and interpretable representation of the yield curve through its level, slope, and curvature factors, its predictive capability is limited under rapidly changing market conditions. In contrast, LSTM models are well suited for learning nonlinear temporal patterns but lack structural interpretability. The proposed DNS-LSTM framework combines the strengths of both approaches, allowing yield curves to be modeled in a structurally consistent yet dynamically flexible manner.

Empirical results indicate that the DNS-LSTM model achieves lower RMSE values than benchmark approaches such as DNS-ARIMA and standalone LSTM models across key maturities, particularly during periods of heightened market volatility. Although the performance gains are incremental rather than substantial, the hybrid model exhibits greater stability and robustness, suggesting that structural

https://doi.org/10.13057/ijas.v9i1.114966
E-ISSN: 2621-086X

econometric models and deep learning methods can complement each other in financial time-series forecasting.

Compared with earlier studies, such as Diebold and Li [5], which established the effectiveness of DNS-based forecasting, and more recent work integrating DNS factors with neural networks [1], this study contributes by maintaining model interpretability while enhancing temporal forecasting performance. In contrast to machine learning approaches applied directly to yield data [16], the proposed framework mitigates overfitting risks by preserving the structural constraints of the DNS model. Overall, the findings support the value of hybrid modeling strategies for yield curve forecasting in both academic research and practical applications.

4. CONCLUSION

This study develops a hybrid dynamic Nelson-Siegel-long short-term memory (DNS-LSTM) framework for forecasting the U.S. Treasury yield curve. By integrating DNS-based yield curve parameterization with LSTM forecasting, the proposed model achieves accurate and stable yield predictions across maturities and consistently outperforms the DNS-ARIMA benchmark. The six-month-ahead projections indicate an upward-sloping yield curve accompanied by declining yield levels, reflecting market expectations of easing monetary conditions.

Despite these promising results, several limitations should be acknowledged. The analysis focuses on a limited set of maturities, which may not fully capture the entire term structure, and the performance of the LSTM component remains sensitive to hyperparameter selection. In addition, attempts to incorporate macroeconomic variables did not improve forecasting accuracy, suggesting the need for more refined feature selection strategies.

Future research may extend this framework by incorporating a broader range of maturities, exploring alternative deep learning architectures such as Transformer-based models, and applying the DNS-LSTM approach to other bond markets, including emerging economies. These extensions would further assess the generalizability and robustness of hybrid yield curve forecasting models.

REFERENCES

[1] Z. Li, B. Wang and Y. Chen, "Incorporating economic indicators and market sentiment effect into US Treasury bond yield prediction with machine learning," *J. Infrastructure, Policy Dev.*, vol. 8, no. 9, p. 7671, 2024, doi: https://doi.org/10.24294/jipd.v8i9.7671.

[2] S. B. Hager, "A global bond: explaining the safe-haven status of US Treasury securities," *Eur. J. Int. Relations*, vol. 23, no. 3, pp. 557–580, 2017, doi: https://doi.org/10.1177/1354066116657400.

[3] D. Cooper, J. Fuhrer, and G. Olivei, "Predicting recessions using the yield curve: the role of the stance of monetary policy," *Fed. Reserv. Bank Bost.*, pp. 1–18, 2020, [Online]. Available: https://www.bostonfed.org/publications/current-policy-perspectives/2020/predicting-recessions-using-the-yield-curve.aspx

[4] C. R. Nelson and A. F. Siegel, "Parsimonious modeling of yield curves," *The Journal of Business*, vol. 60, no. 4. p. 473, 1987. doi: https://doi.org/10.1086/296409.

[5] F. X. Diebold and C. Li, "Forecasting the term structure of government bond yields," *J. Econom.*, vol. 130, pp. 337–364, 2006. doi: https://doi.org/10.1016/j.jeconom.2005.03.005.

[6] S. Fatimah, "Implementasi model dinamik Nelson Siegel pada peramalan yield to maturity," *Fak. Mat. Dan Ilmu Pengetah. Alam Univ. Indones.*, no. January, 2014, doi: https://doi.org/10.13140/RG.2.2.24123.21281.

https://doi.org/10.13057/ijas.v9i1.114966
E-ISSN: 2621-086X

[7] F. X. Diebold, G. D. Rudebusch, and B. Aruoba, "The macroeconomy and the yield curve: a dynamic latent factor approach," *J. Econom.*, vol. 131, pp. 309–338, 2006. doi: https://doi.org/10.1016/j.jeconom.2005.01.011.

[8] D. C. Montgomery, E. A. Peck, and G. G. Vining, "*Introduction to Linear Regression Analysis*," 3rd ed., Canada: John Wiley & Sons, Inc, 2021.

[9] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *"Time Series Analysis: Forecasting and Control,"* 5th ed. Hoboken, NJ, USA: John Wiley & Sons, 2015.

[10] R. J. Hyndman and G. Athanasopoulos, "*Forecasting: Principles and Practice*," 3rd ed. Melbourne: OTEXT, 2021. [Online]. Available: https://otexts.com/fpp3/

[11] S. Shahi, F. H. Fenton, and E. M. Cherry, "Prediction of chaotic time series using recurrent neural networks and reservoir computing techniques: a comparative study," *Mach. Learn. with Appl.*, vol. 8, no. December 2021, p. 100300, 2022, doi: https://doi.org/10.1016/j.mlwa.2022.100300.

[12] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "*Dive into Deep Learning*," Cambridge: Cambridge University Press, 2023. [Online]. Available: https://d2l.ai/chapter_recurrent-modern/lstm.html

[13] M. Kuhn and K. Johnson, "*Applied Predictive Modeling with Applications in R*," vol. 26. 2013. [Online]. Available: http://appliedpredictivemodeling.com/s/Applied_Predictive_Modeling_in_R.pdf

[14] Izzuddin, "Analisis model LSTM dalam peramalan sistem dinamis chaotic persamaan diferensial keuangan," *Repository.Uinjkt.Ac.Id*, 2023, [Online]. Available: https://repository.uinjkt.ac.id/dspace/handle/123456789/66776.

[15] M. Choudhry, *Analysing and Interpreting the Yield Curve*, Second Ed. Chichester: John Wiley & Sons, Inc, 2019.

[16] R. Ronaldo and R. Y. Siregar, "Forecasting Indonesia's Government Bond: machine learning & dynamic Nelson Siegel (DNS) model," vol. 14, no. 2, pp. 109–117, 2017, [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4923808.

https://doi.org/10.13057/ijas.v9i1.114966
E-ISSN: 2621-086X