CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (1) (2026), Pages 1381-1397
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Rolling-Origin Evaluation of Lag-Based Regularized Regression Models for Indonesian Inflation Forecasting

Rehan Risqi Saputra[1*], Atik Wintarti[2], and Riska Wahyu Romadhonia[1]

[1]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Surabaya, State University of Surabaya, Indonesia*
[2]*Department of Data Science, Faculty of Mathematics and Natural Sciences, Surabaya, State University of Surabaya, Indonesia*

Abstract

Inflation forecasting studies often rely on external macroeconomic predictors or complex machine learning models, while the predictive value and limitations of the internal lag structure of Indonesian inflation remain less explicitly examined. This study evaluates Indonesian inflation forecasting within a univariate lag-based framework using monthly inflation data from January 2010 to December 2025. Lagged inflation values are used to construct regression predictors, and OLS, Ridge, LASSO, and Elastic Net are evaluated through rolling-origin forecasting with an expanding window. To strengthen the time-series basis of the analysis, this study also conducts stationarity diagnostics, ACF and PACF analysis, seasonal diagnostics, lag-specification robustness checks, and comparisons with standard forecasting benchmarks, including Naive, Seasonal Naive, AR selected by AIC, and ARIMA selected by AIC. The ADF test produces a p-value of 0.608360, indicating weak evidence of stationarity in level form. Among the regularised regression models, Ridge produces the lowest descriptive forecast errors, with RMSE of 0.405018 and MAE of 0.308967. However, Diebold–Mariano tests indicate that the differences among OLS, Ridge, LASSO, and Elastic Net are not statistically significant. Benchmark comparisons show that the Naive forecast achieves the lowest RMSE of 0.373819, while ARIMA selected by AIC achieves the lowest MAE of 0.279641 and MAPE of 16.560958. Robustness checks also show that the twelve-lag specification is competitive for OLS and Ridge, but it is not uniformly optimal across all models. These findings suggest that the main value of lag-based regularised regression lies in clarifying the limited but useful short-run predictive information contained in the internal temporal structure of Indonesian inflation, rather than in providing a statistically dominant or complete inflation forecasting model.

Keywords: ARIMA Benchmark; Indonesian Inflation; Lagged Predictors; Regularised Regression; Rolling-origin Evaluation; Time Series Forecasting.

1. Introduction

Inflation is one of the most important macroeconomic indicators because its movement affects purchasing power, household expectations, business activity, and the direction of monetary policy. In Indonesia, inflation plays a central role in monetary policy because price stability is

[*]**Corresponding author**. E-mail: rehanrisqisaputra@gmail.com

Submitted: April 17, 2026 Reviewed: May 02, 2026 Accepted: May 27, 2026
DOI: https://doi.org/10.18860/cauchy.v11i1.41964

reflected in low and stable inflation under the monetary policy framework of Bank Indonesia [1]. Bank Indonesia also continuously monitors inflation developments in its policy communication because inflation provides important information for assessing price stability and the broader policy mix [2, 3]. When inflationary pressures arise from supply disruptions, commodity price shocks, or changes in expectations, reliable inflation forecasts become increasingly important. Therefore, inflation forecasting remains a relevant topic in applied statistics, time-series analysis, and quantitative economic research.

Inflation forecasting studies have increasingly incorporated machine learning methods alongside conventional econometric approaches. Machine learning models have been reported to provide competitive forecasting performance in several inflation forecasting settings [4, 5]. In data-rich environments, machine learning methods may improve forecast accuracy by exploiting information from large sets of predictors [6–8]. Other recent studies have explored more flexible approaches, including interpretable machine learning, recurrent neural networks, and advanced time-series filtering or deep learning models [9–11]. Recent review evidence also indicates that machine learning-based inflation forecasting has become an active research area across different horizons and modelling strategies [12]. These studies show the growing relevance of flexible predictive methods for inflation forecasting.

However, the literature also shows that inflation forecasting is not limited to large predictor sets or complex nonlinear models. Time-series approaches remain important because past inflation may contain useful information about future inflation movements. Recent studies continue to examine inflation forecasting through autoregressive and univariate time-series models, including autoregressive distributed lag structures and comparisons between univariate and multivariate forecasting models [13, 14]. This is also related to inflation persistence, where current inflation may be connected to previous inflation dynamics [15]. Therefore, analysing the information contained in lagged inflation values remains useful, especially as a transparent benchmark before more complex models are considered.

Previous studies have explored inflation forecasting using different data structures and modelling strategies. Some studies use large macroeconomic panels or data-rich environments [6, 16]. Other studies use survey expectations, mixed-frequency information, or regularised vector autoregressive models [17, 18]. Country-specific evidence has also been reported for Brazil, China, and other empirical settings [16, 19, 20]. These studies provide valuable evidence, but their focus differs from a setting where inflation is forecast using only its own lagged values. In such a setting, the main question is not whether many external predictors can improve forecasting, but how much predictive information is contained in the internal lag structure of inflation itself.

In the Indonesian context, recent work using Ridge, LASSO, and Elastic Net has mainly examined inflation by using external macroeconomic explanatory variables [21]. This approach is useful for analysing inflation determinants, but it differs from a lag-based forecasting design that uses the inflation series itself as the source of predictors. This distinction is important for two reasons. First, a lag-based model can serve as a simple and reproducible forecasting benchmark because it only requires historical inflation data. Second, lagged inflation values may reflect temporal dependence, inflation persistence, and possible seasonal patterns in monthly inflation data. For practical forecasting, such a model can be useful when external macroeconomic variables are unavailable, delayed, revised, or measured at different frequencies.

The evaluation framework is also important in time-series forecasting. Random train-test splitting is generally inappropriate for time-series data because it breaks the chronological order of observations and may lead to overly optimistic performance estimates [22, 23]. Forecast evaluation is also central in applied inflation forecasting because different institutions and models may produce forecasts with varying accuracy across periods [24]. Forecasting evaluation should therefore estimate models using past observations and assess them on future observations as the forecast origin moves forward over time [22, 23]. Rolling-origin evaluation with an expanding window is consistent with this principle and provides a more realistic assessment of one-step-ahead inflation forecasting performance.

Although the use of lagged inflation values is rooted in autoregressive time-series modelling, this study does not position the lag-based specification as a new forecasting method or as a replacement for standard time-series models. Instead, the study uses the lag-based framework to examine how much predictive information is contained in the internal temporal structure of Indonesian inflation. This framing is important because a univariate lag-based model is simple, transparent, and easy to reproduce, but its usefulness must be evaluated against both regularised regression alternatives and standard time-series forecasting benchmarks.

The contribution of this study is therefore threefold. First, it provides a focused empirical assessment of Indonesian inflation forecasting using only lagged inflation values, allowing the internal structure of the inflation series to be evaluated without relying on external macroeconomic predictors. Second, it compares OLS and regularised lag-based regression models with standard time-series forecasting benchmarks, including naive, seasonal naive, autoregressive, and ARIMA-based forecasts, under the same rolling-origin evaluation framework. Third, it complements forecast accuracy evaluation with time-series diagnostics, lag-specification robustness checks, coefficient stability analysis, and feature selection frequency. This combination allows the study to assess not only which models produce lower forecast errors, but also how the lag structure should be interpreted.

The new insight offered by this study is not the use of regularised regression itself, but the empirical clarification of how far the internal lag structure of Indonesian inflation can support forecasting when it is evaluated against OLS and standard time-series benchmarks under the same rolling-origin design. The results show that Indonesia's inflation history contains useful short-run predictive information, but this information is limited and does not consistently outperform simple benchmark forecasts. This finding positions lag-based regularised regression as a transparent diagnostic tool for examining temporal dependence and coefficient stability, rather than as a stand-alone or complete inflation forecasting model.

Based on these considerations, this study addresses the following research questions:

1. How informative is the internal lag structure of Indonesian inflation for one-step-ahead forecasting under rolling-origin evaluation?
2. How do lag-based regularised regression models compare with OLS and standard time-series forecasting benchmarks?
3. How sensitive are the forecasting results to alternative lag specifications?
4. Which inflation lags are most consistently retained or influential, and how should they be interpreted in light of the time-series diagnostics?

Accordingly, the objective of this study is to evaluate Indonesian inflation forecasting using a univariate lag-based framework, compare regularised regression models with OLS and standard time-series benchmarks, and examine the stability and interpretation of lagged inflation predictors under rolling-origin evaluation.

2. Methods

This section describes the methodological framework used to evaluate Indonesian inflation forecasting under a univariate lag-based design. The explanation begins with the data source and construction of lagged predictors, followed by the forecasting models, diagnostic procedures, rolling-origin evaluation scheme, benchmark comparisons, and coefficient stability analysis.

2.1. Data and Lag-Based Predictor Construction

This study uses monthly Indonesian inflation data obtained from the official Bank Indonesia statistics database [25], covering the period from January 2010 to December 2025. Let y_t denote the inflation rate observed at month t. To transform the univariate inflation series into a regression-based forecasting problem, lagged inflation values are used as predictors. For each month t, the predictor vector is defined as

$$\mathbf{x}_t = (y_{t-1}, y_{t-2}, \dots, y_{t-12})^\top.$$

Thus, each observation consists of twelve lag predictors, namely lag 1 to lag 12, while the response variable is the current inflation value y_t. Because the first twelve observations are required for lag construction, the effective modelling sample begins in January 2011. The use of twelve lags follows the monthly frequency of the data, so that one full annual cycle is represented in the predictor set. This choice is not treated as an optimal lag order, but as a fixed forecasting design intended to represent short-run dependence and possible annual patterns in monthly inflation.

2.2. Lagged Regression Model

The lag-based specification used in this study is closely related to autoregressive time-series modelling, where current inflation is represented as a function of its own past values [13, 14, 23]. The baseline regression model can be written as

$$y_t = \beta_0 + \sum_{j=1}^{12} \beta_j y_{t-j} + \varepsilon_t,$$

where β_0 is the intercept, β_j is the coefficient associated with lag j, and ε_t is the error term. In matrix notation, this model can be expressed as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{y} \in \mathbb{R}^n$ denotes the response vector, $\mathbf{X} \in \mathbb{R}^{n\times p}$ denotes the design matrix of lag predictors, and $p = 12$.

Ordinary least squares is used as an empirical baseline and is defined by

$$\hat{\boldsymbol{\beta}}_{\text{OLS}} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2.$$

The OLS baseline is included because it provides a direct comparison for assessing whether regularised regression improves forecasting performance under a highly correlated lag structure.

2.3. Time-Series Diagnostics

Before model estimation, several time-series diagnostics are conducted to examine the empirical properties of the inflation series. These diagnostics are commonly used to assess temporal dependence, stationarity, and seasonal patterns in time-series forecasting [23]. First, the time-series plot is used to inspect the general movement and volatility of Indonesian inflation over the full sample. Second, the Augmented Dickey–Fuller test is applied to assess whether the inflation series provides evidence of stationarity in level form. Third, the autocorrelation function and partial autocorrelation function are examined up to 36 lags to evaluate the dependence structure of the series and to support the interpretation of lagged inflation predictors.

In addition, a 12-month rolling mean and rolling standard deviation are computed to examine changes in the local mean and volatility of inflation over time. Seasonal diagnostics are also conducted through additive seasonal decomposition with a 12-month period and through the average inflation value for each calendar month. These diagnostics are used to assess whether a simple monthly seasonal pattern is evident in the data. The diagnostics are interpreted as supporting evidence for the lag-based forecasting design, rather than as formal proof that a particular lag order is optimal.

The correlation matrix and variance inflation factor are then used to assess the degree of dependence among lagged predictors. Since lagged predictors are constructed from the same inflation series, high correlation among predictors is expected and is interpreted as part of the autoregressive structure of the data, rather than only as a classical multicollinearity problem [26]. Together with the robustness check, these diagnostics support a cautious interpretation of the twelve-lag design as an annual-cycle specification for monthly data.

2.4. Rolling-Origin Forecasting with Expanding Window

Model evaluation is conducted using rolling-origin forecasting with an expanding window. This design preserves the chronological order of the data and ensures that each forecast is generated only from information available up to the corresponding forecast origin [22, 23]. Let n_0 denote the initial training size. At rolling step h, the training set consists of observations $1, 2, \ldots, n_0 + h - 1$, and the next observation $n_0 + h$ is used as the test point. The one-step-ahead forecast can be written as

$$\hat{y}_{n_0+h} = f_h(\mathbf{x}_{n_0+h}),$$

where f_h denotes the model estimated using the available training sample at rolling step h.

After lag construction, the effective sample contains 180 observations from January 2011 to December 2025. The initial training size is set to 60 observations, corresponding to January 2011 until December 2015. The remaining 120 observations, from January 2016 until December 2025, are used for out-of-sample evaluation through repeated one-step-ahead rolling forecasts. OLS, Ridge, LASSO, and Elastic Net are evaluated under the same rolling-origin design.

2.5. Lag Specification Robustness Check

To provide empirical support for the lag specification, a robustness check is conducted using alternative maximum lag lengths, since predictive information in time-series forecasting may be concentrated in different lag structures [23]. In addition to the main twelve-lag specification, models with maximum lag lengths of 3, 6, and 9 are evaluated to represent short, medium, and longer lag structures in monthly inflation forecasting.

For each lag specification, OLS, Ridge, LASSO, and Elastic Net are estimated using the same rolling-origin evaluation framework. To ensure comparability, all lag specifications are evaluated over the same out-of-sample period from January 2016 to December 2025, producing 120 one-step-ahead forecasts for each model. The twelve-lag design is therefore interpreted as a full annual-cycle forecasting specification, not as an empirically optimal lag order.

2.6. Forecasting Benchmarks

In addition to OLS and regularised regression models, several standard time-series forecasting benchmarks are included to place the proposed lag-based framework within a broader forecasting context. Naive, seasonal naive, autoregressive, and ARIMA-based forecasts are commonly used as reference methods in time-series forecasting [23].

The first benchmark is the naive forecast, which uses the most recent observed inflation value as the one-step-ahead forecast:

$$\hat{y}_t = y_{t-1}.$$

The second benchmark is the seasonal naive forecast, which uses the inflation value from the same month in the previous year:

$$\hat{y}_t = y_{t-12}.$$

The third benchmark is an autoregressive model selected by the Akaike information criterion, which is commonly used for model selection in time-series forecasting [23]. At each rolling step, AR models with lag orders from 1 to 12 are estimated on the training sample, and the model with the lowest AIC is used to generate the one-step-ahead forecast. The fourth benchmark is an ARIMA model selected by AIC from a predefined candidate set. ARIMA models are standard statistical benchmarks for univariate time-series forecasting [23]. The candidate orders are $(1,0,0)$, $(2,0,0)$, $(3,0,0)$, $(1,0,1)$, $(2,0,1)$, $(1,1,0)$, and $(1,1,1)$. At each rolling step, the candidate model with the lowest AIC is selected and used to generate the one-step-ahead forecast.

2.7. Standardisation

Since regularised regression is sensitive to the scale of predictors, standardisation is applied before model estimation [27–29]. At each rolling step, the lag predictors in the training set are standardised using the training mean and training standard deviation. The same transformation is then applied to the corresponding test observation using only the statistics computed from the training set. This procedure avoids information leakage from the test point into the training process. Let x_{tj} denote the value of predictor j at time t. The standardised predictor is defined as

$$z_{tj} = \frac{x_{tj} - \mu_j}{s_j},$$

where μ_j and s_j denote the mean and standard deviation of predictor j in the training set at the corresponding rolling step. The response variable is not standardised.

2.8. Regularised Regression Models

Regularised regression is used to stabilise coefficient estimation when lag predictors are highly correlated. Ridge regression introduces an L_2 penalty and is defined as

$$\hat{\boldsymbol{\beta}}_{\text{Ridge}}(\lambda) = \arg\min_{\boldsymbol{\beta}} \left\{ \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_2^2 \right\}, \quad \lambda \geq 0.$$

LASSO introduces an L_1 penalty and is defined as

$$\hat{\boldsymbol{\beta}}_{\text{LASSO}}(\lambda) = \arg\min_{\boldsymbol{\beta}} \left\{ \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 \right\}, \quad \lambda \geq 0.$$

Because the L_1 penalty is not differentiable at zero, the LASSO solution is obtained numerically rather than through a closed-form expression [28]. Elastic Net combines the L_1 and L_2 penalties and is defined as

$$\hat{\boldsymbol{\beta}}_{\text{EN}}(\lambda, \alpha) = \arg\min_{\boldsymbol{\beta}} \left\{ \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \left(\frac{1-\alpha}{2}\|\boldsymbol{\beta}\|_2^2 + \alpha\|\boldsymbol{\beta}\|_1 \right) \right\}, \quad 0 \leq \alpha \leq 1.$$

Here, λ controls the overall penalty strength, while α determines the balance between the L_1 and L_2 penalties. Ridge is commonly used to stabilise coefficient estimates when predictive information is distributed across correlated predictors [27]. LASSO encourages sparsity through variable selection [28]. Elastic Net combines shrinkage and selection in the presence of correlated predictors [29].

2.9. Hyperparameter Tuning

The regularisation parameters are selected through time-series cross-validation within each rolling training sample. This inner validation procedure preserves chronological order and is consistent with forecasting evaluation for dependent observations [23]. For Ridge and LASSO, the tuning parameter is the penalty value λ. For Elastic Net, both λ and α are tuned.

The penalty grid consists of 40 logarithmically spaced values from 10^{-4} to 10^2. For Elastic Net, the candidate α values are 0.1, 0.3, 0.5, 0.7, and 0.9. The inner time-series cross-validation uses five splits. The same hyperparameter search space is applied at every rolling step. The hyperparameter combination that minimises the average validation mean squared error is selected. After the optimal hyperparameters are obtained, the model is re-estimated on the full training set at the corresponding rolling step and then used to generate the one-step-ahead forecast.

2.10. Evaluation Metrics and Forecast Comparison

Forecast accuracy is assessed over the full set of rolling forecasts using RMSE, MAE, MAPE, and R^2 [23]. Let $\{y_t\}_{t=1}^m$ denote the realised values on the rolling test set and let $\{\hat{y}_t\}_{t=1}^m$ denote the corresponding forecasts. The root mean squared error is defined as

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{t=1}^{m}(y_t - \hat{y}_t)^2}. \tag{1}$$

The mean absolute error is defined as

$$\text{MAE} = \frac{1}{m}\sum_{t=1}^{m}|y_t - \hat{y}_t|. \tag{2}$$

The mean absolute percentage error is defined as

$$\text{MAPE} = \frac{100}{m}\sum_{t=1}^{m}\left|\frac{y_t - \hat{y}_t}{y_t}\right|. \tag{3}$$

The coefficient of determination is defined as

$$R^2 = 1 - \frac{\sum_{t=1}^{m}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{m}(y_t - \bar{y})^2}, \quad \bar{y} = \frac{1}{m}\sum_{t=1}^{m} y_t. \tag{4}$$

RMSE and MAE are treated as the main error measures because they are expressed in the same scale as the inflation data. MAPE is reported as a supplementary percentage-based measure. Since inflation values may be close to zero, MAPE can become unstable and is therefore interpreted cautiously. The out-of-sample R^2 is also interpreted as a descriptive measure of predictive fit over the rolling test period.

To complement the descriptive accuracy measures, the Diebold–Mariano test is used to compare forecast accuracy between pairs of models [30]. Because the test compares two competing forecasts at a time, it is applied pairwise to the out-of-sample forecast errors of OLS, Ridge, LASSO, and Elastic Net using squared error loss. For each pair of competing forecasts, the loss differential is defined as

$$d_t = e_{1t}^2 - e_{2t}^2,$$

where $e_{1t} = y_t - \hat{y}_{1t}$ and $e_{2t} = y_t - \hat{y}_{2t}$ denote the forecast errors from the first and second models, respectively. The null hypothesis states that the two competing models have equal predictive accuracy, or equivalently that the expected loss differential is zero. A statistically significant result indicates that the difference in forecast accuracy between the two models is unlikely to be due to sampling variation alone.

2.11. Coefficient Stability and Feature Selection Frequency

In addition to forecast accuracy, this study evaluates coefficient stability across rolling steps. This analysis is important because predictive performance alone does not fully explain how consistently the lag structure contributes over time. Coefficient stability is summarised using the mean and standard deviation of each lag coefficient across rolling steps for OLS, Ridge, LASSO, and Elastic Net.

Since the lag predictors are standardised separately within each rolling training window, the reported coefficients are interpreted on the standardised predictor scale. Therefore, the coefficient stability results describe the relative stability of lag contributions within the rolling forecasting design, rather than original-scale economic effects. This limitation is considered when interpreting the magnitude of the coefficients.

For LASSO and Elastic Net, feature selection frequency is also examined. Feature selection frequency is defined as the proportion of rolling steps in which a lag receives a non-zero coefficient. This measure indicates how consistently a given lag is retained by the regularised model throughout the rolling forecasting process. The use of non-zero coefficient frequency is relevant for LASSO and Elastic Net because the L_1 penalty can shrink some coefficients exactly to zero, enabling variable selection [28, 29]. Ridge is not evaluated using non-zero selection frequency because its L_2 penalty shrinks coefficients but generally does not set them exactly to zero [27].

2.12. Computational Implementation

All computations are carried out in Python using standard scientific computing libraries. Data preparation, diagnostics, rolling forecasting, benchmark estimation, model tuning, robustness checks, and coefficient stability analysis are implemented consistently under the rolling-origin evaluation design.

3. Results and Discussion

This section presents the empirical findings from the lag-based forecasting evaluation. The discussion begins with time-series diagnostics to clarify the structure of Indonesian inflation before examining lag robustness, forecast accuracy, benchmark comparisons, coefficient stability, and feature selection patterns.

3.1. Time-Series Diagnostic Results

The monthly Indonesian inflation series is presented in Fig. 1. The original dataset contains 192 monthly observations from January 2010 to December 2025. The vertical marker shows the start of the out-of-sample evaluation in January 2016 after the initial training period. The series fluctuates across the sample, with periods of higher volatility and periods of relatively stable inflation. This pattern supports treating the forecasting task as a time-series problem rather than as an ordinary cross-sectional regression problem.

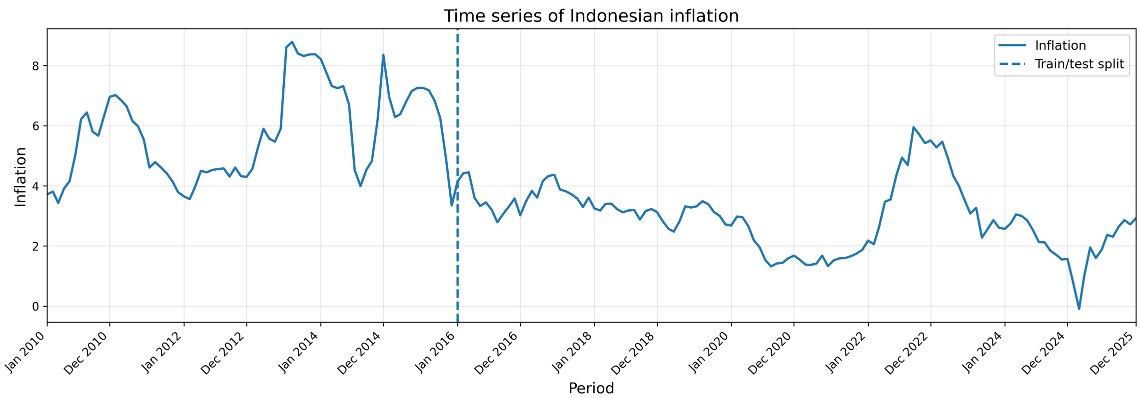

Fig. 1: Time series of Indonesian inflation.

The Augmented Dickey–Fuller test was used to examine stationarity in the inflation series. The test produced an ADF statistic of -1.344962 with a p-value of 0.608360. Therefore, the null hypothesis of a unit root cannot be rejected at conventional significance levels. This result indicates that the inflation series does not provide strong evidence of stationarity in level form over the full sample. For this reason, the forecasting results are interpreted as out-of-sample predictive performance under a rolling-origin design, rather than as evidence from a strictly stationary data-generating process.

The ACF and PACF plots are presented in Fig. 2. The results indicate that Indonesian inflation contains temporal dependence, especially at short lags. This supports the use of lagged

Table 1: Augmented Dickey–Fuller test result

Statistic	Value
ADF statistic	-1.344962
p-value	0.608360
Used lag	12
Number of observations	179

inflation values as predictors. However, the diagnostic evidence does not imply that twelve lags are empirically optimal. Instead, the twelve-lag specification should be interpreted as a full annual-cycle representation for monthly data.

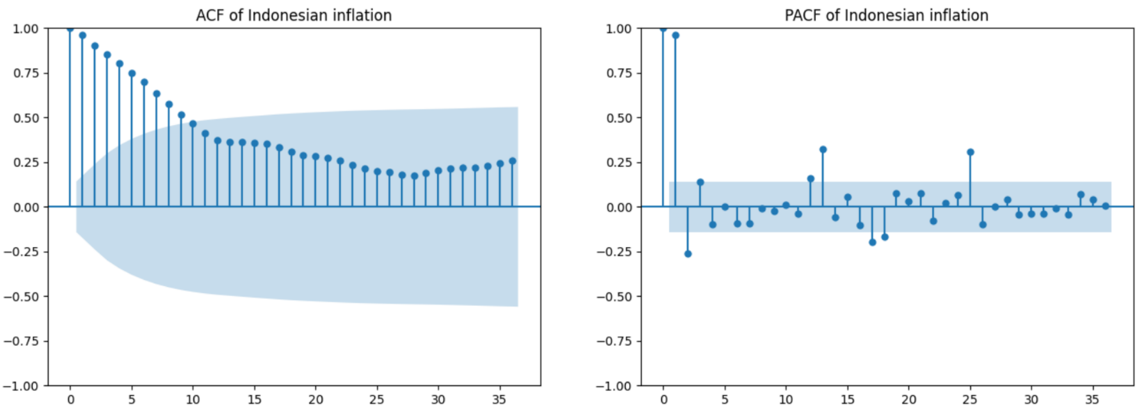

Fig. 2: ACF and PACF of Indonesian inflation.

A 12-month rolling mean and rolling standard deviation were also examined to assess changes in local mean and volatility. The results indicate time-varying inflation dynamics, supporting the use of rolling-origin evaluation across different economic periods.

Seasonal diagnostics were examined using seasonal decomposition and average inflation by calendar month. As reported in Table 2, the monthly averages are relatively close across months, suggesting that strong deterministic monthly seasonality is not evident descriptively. This finding is also consistent with the poor performance of the Seasonal Naive benchmark reported later in this section. Therefore, lag 11 and lag 12 should be interpreted cautiously as possible near-annual statistical signals, not as evidence of a stable annual seasonal pattern.

Table 2: Average inflation by calendar month

Month	Mean	Count	Month	Mean	Count	Month	Mean	Count
1	3.928750	16	5	4.037500	16	9	3.955000	16
2	3.881875	16	6	4.046250	16	10	3.951250	16
3	3.975625	16	7	4.060000	16	11	3.955625	16
4	4.064375	16	8	4.023125	16	12	3.986250	16

After the lag variables are constructed, the correlation structure among lag 1 to lag 12 is presented in Fig. 3. The figure indicates that several lag predictors are strongly correlated. This pattern is expected because all predictors are lagged values from the same inflation series. Therefore, the high correlation among predictors should be interpreted not only as a classical multicollinearity issue, but also as a reflection of the autoregressive dependence structure in monthly inflation [13, 14, 23].

The magnitude of dependence among lag predictors is further examined using the variance inflation factor. The results are reported in Table 3. Most lag predictors have VIF values far above 10. The highest VIF is observed for lag 6, namely 44.329017, followed by lag 5 with 44.128220 and lag 7 with 44.087868. Even the smallest VIF among the lag predictors remains

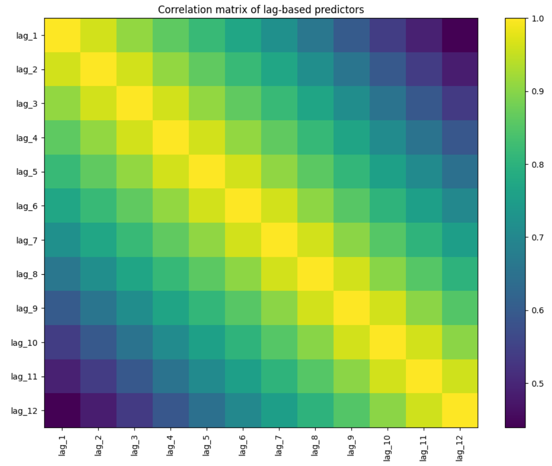

Fig. 3: Correlation matrix of lag-based predictors.

high, namely 14.658757 for lag 12. These results confirm strong dependence among lag predictors, which motivates the use of regularised regression [27, 29]. Since this dependence arises from lagged observations of the same series, it should be interpreted within the time-series forecasting context.

Table 3: Variance inflation factor of lag predictors

Variable	VIF	Variable	VIF	Variable	VIF
lag 1	15.350993	lag 5	44.128220	lag 9	43.266362
lag 2	39.264958	lag 6	44.329017	lag 10	41.468190
lag 3	42.451085	lag 7	44.087868	lag 11	38.254638
lag 4	43.679471	lag 8	43.613946	lag 12	14.658757

3.2. Lag Specification Robustness

To assess whether the twelve-lag specification is empirically supported, a robustness check was conducted using alternative maximum lag lengths of 3, 6, 9, and 12. All specifications were evaluated over the same out-of-sample period from January 2016 to December 2025, producing 120 one-step-ahead forecasts for each model. This ensures that the comparison across lag specifications is based on the same forecasting period.

Table 4: RMSE comparison across alternative lag specifications

Maximum lag	OLS	Ridge	LASSO	Elastic Net
3	0.412557	0.412829	0.415335	0.415893
6	0.406904	0.406662	0.407658	0.407608
9	0.411240	0.410439	0.405405	0.408060
12	0.405858	0.405018	0.412938	0.409658

The robustness results show that the twelve-lag specification is not uniformly optimal across all models. For OLS and Ridge, the twelve-lag specification gives the lowest RMSE. For LASSO, the nine-lag specification gives the lowest RMSE, while for Elastic Net, the six-lag specification gives the lowest RMSE. These findings indicate that the twelve-lag design remains competitive, especially for OLS and Ridge, but it should not be interpreted as the empirically optimal lag order for all models.

Therefore, the twelve-lag specification is retained in the main analysis as a full annual-cycle representation for monthly inflation data. In this study, lag 1 to lag 12 are used to represent short-run to near-annual dependence within a common monthly forecasting design. The robustness check also shows that shorter lag structures contain substantial predictive information, which is consistent with the importance of short-run inflation persistence.

3.3. Forecast Accuracy and Benchmark Comparison

Forecast performance is evaluated using RMSE, MAE, MAPE, and out-of-sample R^2, as defined in Eq. (1), Eq. (2), Eq. (3), and Eq. (4). The results are reported in Table 5. In addition to OLS, Ridge, LASSO, and Elastic Net, standard time-series benchmarks are included, namely Naive, Seasonal Naive, AR selected by AIC, and ARIMA selected by AIC.

Table 5: Out-of-sample forecasting performance with benchmark models

Model	RMSE	MAE	MAPE	R^2
OLS	0.405858	0.310260	18.102234	0.867244
Ridge	0.405018	0.308967	18.086936	0.867792
LASSO	0.412938	0.319403	20.365915	0.862572
Elastic Net	0.409658	0.316468	18.511180	0.864746
Naive	0.373819	0.281417	18.759626	0.887376
Seasonal Naive	1.793507	1.373833	76.257929	-1.592467
AR selected by AIC	0.408411	0.301243	18.545172	0.865568
ARIMA selected by AIC	0.390175	0.279641	16.560958	0.877305

Among the regularised regression models, Ridge produces the lowest descriptive forecast errors, with RMSE of 0.405018, MAE of 0.308967, and MAPE of 18.086936. However, when standard time-series benchmarks are included, Ridge no longer has the lowest descriptive error across all metrics. The Naive benchmark produces the lowest RMSE and the highest out-of-sample R^2, while the ARIMA selected by AIC benchmark produces the lowest MAE and MAPE.

This comparison provides a more balanced interpretation of the lag-based regularised regression framework. Ridge can therefore be described only as the model with the lowest descriptive forecast error within the regularised regression group, not as the statistically superior forecasting model overall. The broader benchmark comparison favours a more cautious interpretation because the Naive and ARIMA selected by AIC benchmarks produce lower descriptive errors on several metrics. The strong performance of the Naive benchmark suggests that short-run inflation persistence is highly informative for one-step-ahead forecasting. The competitive performance of ARIMA also indicates that standard time-series models remain important benchmarks for Indonesian inflation forecasting.

The Seasonal Naive benchmark performs poorly, with RMSE of 1.793507 and negative out-of-sample R^2. This suggests that a simple annual repetition rule is not sufficient to describe the dynamics of Indonesian inflation. Therefore, the use of lag 12 in the main lag-based design should not be interpreted as evidence of strong deterministic seasonality. Instead, it should be understood as an annual-cycle representation whose usefulness must be evaluated empirically.

The AR model selected by AIC most frequently chose lag order 3, while the ARIMA benchmark most frequently selected ARIMA$(1, 1, 1)$. The frequent selection of short AR lag order supports the importance of short-run dependence, while the frequent selection of differenced ARIMA models is consistent with the ADF test, which did not provide strong evidence of stationarity in the level series.

Since the differences among OLS, Ridge, LASSO, and Elastic Net are relatively small, the Diebold–Mariano test is used to compare forecast errors within the lag-based regression group. The test is applied pairwise using squared error loss. The results are presented in Table 7.

The Diebold–Mariano test results show that none of the pairwise differences among OLS, Ridge, LASSO, and Elastic Net are statistically significant at the 5% level. Therefore, although Ridge has the lowest descriptive error within the regularised regression group, its advantage over the other lag-based regression models is not statistically significant. This means that Ridge should be interpreted as a descriptively competitive model within the regularised regression

Table 6: Selected AR lag order and ARIMA order frequencies

Benchmark	Selected order	Frequency
AR selected by AIC	3	94
AR selected by AIC	4	4
AR selected by AIC	7	22
ARIMA selected by AIC	$(1, 1, 1)$	117
ARIMA selected by AIC	$(1, 0, 1)$	3

Table 7: Diebold–Mariano test for pairwise forecast comparison

Model 1	Model 2	DM statistic	p-value	Mean loss difference
OLS	Ridge	0.599772	0.549799	0.000681
OLS	LASSO	-0.778496	0.437822	-0.005797
OLS	Elastic Net	-0.454551	0.650260	-0.003099
Ridge	LASSO	-0.935114	0.351623	-0.006478
Ridge	Elastic Net	-0.605677	0.545883	-0.003780
LASSO	Elastic Net	0.902355	0.368691	0.002698

group, rather than as an empirically dominant forecasting model. This result supports a cautious interpretation of Ridge's descriptive advantage and positions the lag-based regularised framework as a transparent analytical benchmark.

The comparison between actual inflation and predicted inflation in the out-of-sample period is shown in Fig. 4. The models generally follow the broad movement of inflation, but larger deviations appear during periods of sharper fluctuation. This is expected because the models use only past inflation values as predictors. When inflation changes are driven by sudden shocks, policy adjustments, supply disruptions, or commodity price movements, lagged inflation alone may not fully capture the new information immediately. This explains why forecast errors can increase during volatile periods even when the overall error measures remain relatively low.

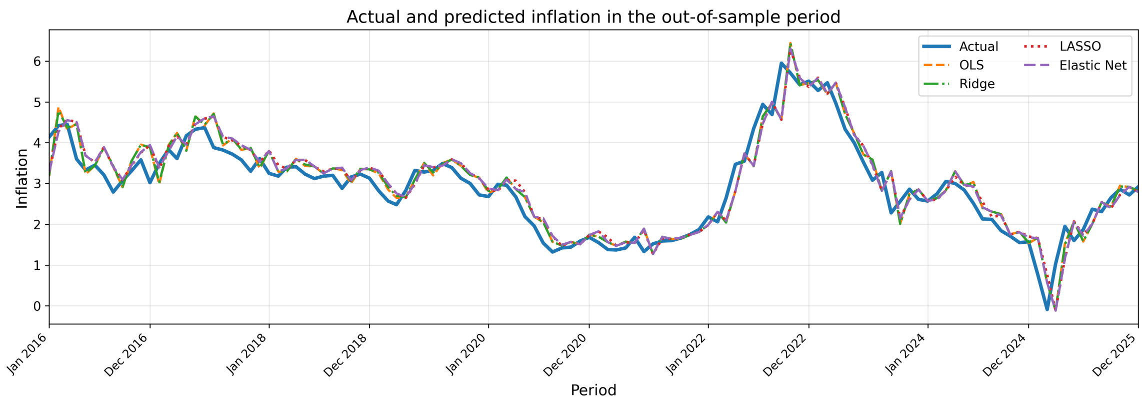

Fig. 4: Actual and predicted inflation in the out-of-sample period.

3.4. Coefficient Stability and Feature Selection Frequency

In addition to forecast accuracy, coefficient stability is examined to understand how consistently each lag contributes across rolling steps. Table 8 reports the mean and standard deviation of the estimated coefficients across rolling forecasts. This analysis complements forecast accuracy because similar predictive performance may arise from different coefficient patterns.

Because the predictors are standardised separately within each rolling training window, the reported coefficient summaries are interpreted on the standardised predictor scale. Therefore, the coefficient means and standard deviations should be understood as indicators of relative stability within the rolling forecasting design, rather than original-scale economic effects. This

Table 8: Coefficient stability summary across rolling steps

Lag	OLS mean	OLS std	Ridge mean	Ridge std	LASSO mean	LASSO std	EN mean	EN std
lag 1	2.247013	0.165593	2.211345	0.198085	1.860530	0.357199	1.895343	0.382920
lag 2	-0.899472	0.062148	-0.843793	0.124698	-0.287177	0.357746	-0.358289	0.400281
lag 3	0.410485	0.059901	0.375238	0.088462	0.092494	0.163139	0.136761	0.182745
lag 4	-0.095336	0.106702	-0.078246	0.108309	-0.010668	0.066785	-0.022515	0.073527
lag 5	0.061626	0.103068	0.057736	0.099193	0.019930	0.052655	0.025942	0.063441
lag 6	0.096310	0.061529	0.092975	0.058597	0.016820	0.041004	0.026711	0.050050
lag 7	-0.159248	0.039165	-0.150693	0.043064	-0.035853	0.058576	-0.050813	0.067230
lag 8	0.045751	0.073045	0.035724	0.075989	-0.001499	0.042480	0.002844	0.050880
lag 9	-0.124962	0.094452	-0.114467	0.096242	-0.025995	0.064136	-0.037957	0.074629
lag 10	0.226274	0.068463	0.211596	0.076069	0.044584	0.091932	0.068399	0.105931
lag 11	-0.424677	0.041778	-0.408137	0.058955	-0.126754	0.157127	-0.169025	0.176521
lag 12	0.262032	0.042258	0.252699	0.050067	0.071960	0.110693	0.100500	0.124107

is particularly important when comparing coefficients across rolling steps, because the training mean and standard deviation may change over time.

The coefficient stability results indicate that lag 1 has the largest average coefficient under all models. This supports the role of short-run inflation persistence, where the most recent inflation observation contains substantial information for the next one-step-ahead forecast [15]. OLS and Ridge show relatively similar coefficient patterns, while Ridge slightly shrinks the coefficients. LASSO and Elastic Net reduce the influence of several lags more strongly. This difference is consistent with the theoretical behaviour of the penalty terms, where Ridge emphasises coefficient shrinkage and LASSO encourages sparse selection [27, 28].

For LASSO and Elastic Net, feature selection frequency is also examined. The corresponding results are reported in Table 9.

Table 9: Selection frequency of lag predictors

Lag	LASSO	Elastic Net	Lag	LASSO	Elastic Net	Lag	LASSO	Elastic Net
lag 1	100.00	100.00	lag 5	31.67	38.33	lag 9	35.00	43.33
lag 2	55.00	54.17	lag 6	19.17	29.17	lag 10	21.67	33.33
lag 3	39.17	48.33	lag 7	39.17	48.33	lag 11	84.17	80.83
lag 4	22.50	30.00	lag 8	58.33	59.17	lag 12	44.17	49.17

The selection frequency results show that lag 1 is selected in all rolling steps by both LASSO and Elastic Net. This confirms the central role of short-run inflation persistence, where the most recent inflation observation carries substantial information for one-step-ahead forecasting. Beyond lag 1, lag 11 and lag 8 are selected relatively frequently, while lag 12 is selected with moderate frequency. These lags should be interpreted as recurring statistical signals in the rolling selection process, not as direct evidence of specific economic or policy cycles.

The relatively frequent selection of lag 11 and the moderate selection of lag 12 point to tentative near-annual dependence in the lag structure. However, the seasonal diagnostics and the poor performance of the Seasonal Naive benchmark do not support a strong deterministic annual pattern. Therefore, these near-annual lags are interpreted as possible temporal-dependence signals rather than formal evidence of seasonality. Lag 8 may reflect medium-range dependence, but its economic interpretation remains limited because the model does not include external macroeconomic variables such as exchange rates, commodity prices, administered prices, or supply-side indicators.

3.5. Discussion

The results show that lagged inflation values contain useful information for one-step-ahead Indonesian inflation forecasting, especially through short-run dependence. This is reflected in the competitive performance of OLS, Ridge, ARIMA-based benchmarks, and the strong performance of the Naive benchmark.

However, regularised regression does not statistically dominate the alternatives. Ridge achieves the lowest descriptive error among the regularised regression models, but the Diebold–Mariano test shows that its differences relative to OLS, LASSO, and Elastic Net are not statistically significant. When standard time-series benchmarks are included, Naive and ARIMA selected by AIC produce lower descriptive errors on several metrics. Therefore, the result should not be read as evidence that Ridge is generally superior for Indonesian inflation forecasting. The value of regularised lag-based regression lies mainly in providing a transparent framework for examining correlated lag predictors and coefficient stability.

The robustness check clarifies that the twelve-lag specification is competitive for OLS and Ridge, but it is not universally optimal. Shorter lag structures, especially six and nine lags, perform better for some sparse models. This supports a more cautious interpretation of the twelve-lag design as an annual-cycle representation for monthly inflation data.

The diagnostics and benchmark results also limit the interpretation of seasonality. Although lag 11 and lag 12 appear in the selection analysis, seasonal decomposition, monthly averages, and the poor performance of the Seasonal Naive benchmark do not support a strong deterministic annual pattern. These near-annual lags are therefore better understood as tentative temporal-dependence signals within the univariate lag structure, not as evidence that Indonesian inflation follows a stable annual seasonal pattern. Similarly, lag 8 may indicate medium-range dependence, but its interpretation remains limited without external macroeconomic variables.

From a practical perspective, the proposed framework should be positioned as a simple and transparent baseline for preliminary inflation monitoring. It only requires historical inflation data and can be updated easily. However, it should not be viewed as a complete inflation forecasting system because it excludes external macroeconomic variables such as exchange rates, commodity prices, administered prices, and supply-side indicators. The framework also does not formally test for structural breaks, so possible changes in inflation dynamics across the 2010–2025 period should be interpreted as a limitation of the study. Overall, univariate lag-based regularised regression is useful as a complementary analytical benchmark, while any claims of superiority should be made cautiously.

4. Conclusion

This study evaluated Indonesian inflation forecasting using a univariate lag-based framework under rolling-origin evaluation with an expanding window. Monthly inflation data from January 2010 to December 2025 were transformed into lagged predictor structures, and OLS, Ridge, LASSO, and Elastic Net were compared with Naive, Seasonal Naive, AR selected by AIC, and ARIMA selected by AIC.

The main contribution of this study is to provide a critical empirical assessment of the forecasting value and limitations of the internal lag structure of Indonesian inflation. Rather than proposing regularised regression as a new forecasting method, this study shows that a univariate lag-based framework can be useful for diagnosing short-run inflation persistence and coefficient stability, while also revealing that such information remains limited when compared with simple time-series benchmarks. Ridge produced the lowest descriptive forecast errors among the regularised regression models, but the Diebold–Mariano test indicates that its differences relative to OLS, LASSO, and Elastic Net are not statistically significant. In addition, the Naive and ARIMA benchmarks produced lower descriptive errors on several accuracy measures. Therefore, regularised lag-based regression should be interpreted as a transparent complementary benchmark for understanding the internal temporal structure of Indonesian inflation, rather than as a statistically dominant or generally superior forecasting model.

The lag specification and coefficient stability results support a cautious interpretation of the model. The twelve-lag specification is competitive for OLS and Ridge, but it is not uniformly optimal across all models. Lag 1 is the most consistently important predictor, supporting short-run inflation persistence. Lag 8, lag 11, and lag 12 also appear in the rolling selection process, but these lags should be interpreted as statistical signals rather than direct evidence of specific economic, policy, or stable seasonal cycles.

From a practical perspective, the proposed framework should be positioned as a simple and transparent baseline for preliminary inflation monitoring, not as an operational forecasting system for policy use. It only requires historical inflation data and can be updated easily, which makes it useful as a reproducible benchmark when external predictors are unavailable or delayed. However, it should not be viewed as a complete inflation forecasting system because it excludes external macroeconomic variables such as exchange rates, commodity prices, administered prices, and supply-side indicators. The framework also does not formally test for structural breaks, so possible changes in inflation dynamics across the 2010–2025 period should be interpreted as a limitation of the study. Overall, univariate lag-based regularised regression is useful as a complementary analytical benchmark, while any claims of superiority should be made cautiously.

CRediT Author Contributions

Rehan Risqi Saputra: Conceptualization, Methodology, Formal Analysis, Data Curation, Software, Visualization, Writing–Original Draft. **Atik Wintarti:** Supervision, Validation, Methodology, Writing–Review & Editing. **Riska Wahyu Romadhonia:** Supervision, Validation, Writing–Review & Editing.

Declaration of Generative AI and AI-assisted Technologies

Generative AI and AI-assisted technologies, specifically ChatGPT, were used during the preparation of this manuscript for language refinement, drafting support, and revision assistance. All analyses, results, interpretations, conclusions, and key scientific substance were developed entirely by the authors and were not generated by AI. The authors reviewed, verified, and approved the final manuscript and take full responsibility for its content.

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgments

This research received no external funding.

Data and Code Availability

The monthly Indonesian inflation data analysed in this study were obtained from the official Bank Indonesia statistics database [25]. The processed dataset and code supporting the findings of this study are available from the corresponding author upon reasonable request.

References

[1] Bank Indonesia. *Moneter*. 2026. https://www.bi.go.id/id/fungsi-utama/moneter/default.aspx.

[2] Bank Indonesia. *Monetary Policy Review March 2025*. 2025. https://www.bi.go.id/en/publikasi/laporan/Pages/TKM-Maret-2025.aspx.

[3] Bank Indonesia. *Monetary Policy Review May 2025*. 2025. https://www.bi.go.id/en/publikasi/laporan/Pages/TKM-Mei-2025.aspx.

[4] P. K. Das and P. K. Das. "Forecasting and Analyzing Predictors of Inflation Rate: Using Machine Learning Approach". In: *Journal of Quantitative Economics* 22.2 (2024), pp. 493–517. DOI: 10.1007/s40953-024-00384-z.

[5] S. Aras and P. J. G. Lisboa. "Explainable Inflation Forecasts by Machine Learning Models". In: *Expert Systems with Applications* 207 (2022), p. 117982. DOI: 10.1016/j.eswa.2022.117982.

[6] Marcelo C. Medeiros, Gabriel F. R. Vasconcelos, Álvaro Veiga, and Eduardo Zilberman. "Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods". In: *Journal of Business and Economic Statistics* 39.1 (2021), pp. 98–119. DOI: 10.1080/07350015.2019.1637745.

[7] A. A. Naghi, E. O'Neill, and M. Danielova Zaharieva. "The Benefits of Forecasting Inflation with Machine Learning: New Evidence". In: *Journal of Applied Econometrics* 39.7 (2024), pp. 1321–1331. DOI: 10.1002/jae.3088.

[8] N. Anesti, E. Kalamara, and G. Kapetanios. "Forecasting with Machine Learning Methods and Multiple Large Datasets". In: *Econometrics and Statistics* (2024). DOI: 10.1016/j.ecosta.2024.08.003.

[9] A. Ben Mrad and H. M. Alsowayyan. "Interpretable Machine Learning for Price Index Forecasting: A Case Study with Rolling Windows and SHAP". In: *Engineering, Technology and Applied Science Research* 16.1 (2026), pp. 30954–30962. DOI: 10.48084/etasr.14863.

[10] Livia Paranhos. "Predicting Inflation with Recurrent Neural Networks". In: *International Journal of Forecasting* (2025). DOI: 10.1016/j.ijforecast.2024.07.010.

[11] A. J. Martínez Casares. "Analysis of Improvements in Inflation Prediction Performance through the Use of Hybrid Filter Models and Many-to-one Neural Networks". In: *Computational Economics* (2026). DOI: 10.1007/s10614-025-11220-8.

[12] H. Slimani and C. Chourabi. "Modeling Inflation with Machine Learning: A Cross-Horizon Systematic Review". In: *International Journal of Data Science and Analytics* 21 (2026), p. 65. DOI: 10.1007/s41060-025-00905-w.

[13] Regi Muzio Ponziani. "Inflation Forecasting Using Autoregressive Distributed Lag (ARDL) Models". In: *Jurnal Ekonomi & Studi Pembangunan* 24.2 (2023), pp. 316–330. DOI: 10.18196/jesp.v24i2.17620.

[14] Hasnain Iftikhar, Faridoon Khan, Paulo Canas Rodrigues, Abdulmajeed Atiah Alharbi, and Jeza Allohibi. "Forecasting of Inflation Based on Univariate and Multivariate Time Series Models: An Empirical Application". In: *Mathematics* 13.7 (2025), p. 1121. DOI: 10.3390/math13071121.

[15] Alessandra Canepa. "Inflation Dynamics and Persistence: The Importance of the Uncertainty Channel". In: *The North American Journal of Economics and Finance* 72 (2024), p. 102135. DOI: 10.1016/j.najef.2024.102135.

[16] N. Huang, Y. Qi, and J. Xia. "China's Inflation Forecasting in a Data-Rich Environment: Based on Machine Learning Algorithms". In: *Applied Economics* 57.17 (2025), pp. 1995–2020. DOI: 10.1080/00036846.2024.2322572.

[17] F. Huber, L. Onorante, and M. Pfarrhofer. "Forecasting Euro Area Inflation Using a Huge Panel of Survey Expectations". In: *International Journal of Forecasting* 40.3 (2024), pp. 1042–1054. DOI: 10.1016/j.ijforecast.2023.09.003.

[18] T. Aliaj, M. Ciganovic, and M. Tancioni. "Nowcasting Inflation with Lasso-Regularized Vector Autoregressions and Mixed Frequency Data". In: *Journal of Forecasting* 42.3 (2023), pp. 464–480. DOI: 10.1002/for.2944.

[19] G. S. Araújo and W. P. Gaglianone. "Machine Learning Methods for Inflation Forecasting in Brazil: New Contenders versus Classical Models". In: *Latin American Journal of Central Banking* 4 (2023), p. 100087. DOI: 10.1016/j.lajcb.2023.100087.

[20] B. Oancea, M. Simionescu, and R. Pospisil. "Do Machine Learning Techniques Outperform Autoregressive Distributed Lag Models in Inflation Forecasting?" In: *Prague Economic Papers* 34.4 (2025), pp. 495–558. DOI: 10.18267/j.pep.898.

[21] Husna Afanyn Khoirunissa, Andreas Rony Wijaya, Bayutama Isnaini, and Kiki Ferawati. "Analisis Faktor-Faktor Penyebab Inflasi di Indonesia Menggunakan Regresi Ridge, LASSO, dan Elastic-Net". In: *Indonesian Journal of Applied Statistics* 7.2 (2024), pp. 121–130. DOI: 10.13057/ijas.v7i2.96921.

[22] Leonard J. Tashman. "Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review". In: *International Journal of Forecasting* 16.4 (2000), pp. 437–450. DOI: 10.1016/S0169-2070(00)00065-0.

[23] Rob J. Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. 3rd ed. Available: https://otexts.com/fpp3/. Melbourne: OTexts, 2021.

[24] George S. Tavlas, Yan Wang, Eleni Argiri, Stephen G. Hall, Anna Momtsia, and Ifigeneia Skotida. "An Evaluation of the Inflation Forecasting Performance of the European Central Bank, the Federal Reserve, and the Bank of England". In: *Journal of Forecasting* 43.4 (2024), pp. 932–947. DOI: 10.1002/for.3063.

[25] Bank Indonesia. *Data Inflasi*. 2026. https://www.bi.go.id/id/statistik/indikator/data-inflasi.aspx.

[26] N. Herawati, A. Wijayanti, A. Sutrisno, Nusyirwan, and Misgiyati. "The Performance of Ridge Regression, LASSO, and Elastic-Net in Controlling Multicollinearity: A Simulation and Application". In: *Journal of Modern Applied Statistical Methods* 23.2 (2024), pp. 1–13. DOI: 10.56801/JMASM.V23.I2.4.

[27] Arthur E. Hoerl and Robert W. Kennard. "Ridge Regression: Biased Estimation for Nonorthogonal Problems". In: *Technometrics* 12.1 (1970), pp. 55–67. DOI: 10.1080/00401706.1970.10488634.

[28] Robert Tibshirani. "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society: Series B* 58.1 (1996), pp. 267–288. DOI: 10.1111/j.2517-6161.1996.tb02080.x.

[29] Hui Zou and Trevor Hastie. "Regularization and Variable Selection via the Elastic Net". In: *Journal of the Royal Statistical Society: Series B* 67.2 (2005), pp. 301–320. DOI: 10.1111/j.1467-9868.2005.00503.x.

[30] Francis X. Diebold and Roberto S. Mariano. "Comparing Predictive Accuracy". In: *Journal of Business & Economic Statistics* 13.3 (1995), pp. 253–263. DOI: 10.1080/07350015.1995.10524599.