ANALYSIS OF INDONESIAN INFLATION 2006–2024 USING PYTHON LIBRARY: ARIMA, SARIMA, AND EXPONENTIAL SMOOTHING MODELS

MOHAMMAD NASRI AW
Sekolah Tinggi Ilmu Ekonomi Indonesia Malang
Email: nasriaw@gmail.com

SUDARJO
Sekolah Tinggi Ilmu Ekonomi Indonesia Malang
Email: sudarjokusumo@gmail.com

ABSTRACT

This paper conducts an in-depth analysis of Indonesia's monthly inflation rate from January 2006 to December 2024, employing advanced time series techniques to uncover underlying patterns and to develop a robust predictive framework. Utilizing Python's TimeSeriesSplit for cross-validation, we implement and compare multiple forecasting models—specifically ARIMA, SARIMA, and Exponential Smoothing—evaluating their performance across a rolling forecast horizon. The study identifies key periods of volatility linked to the 2008 global financial crisis, domestic fuel subsidy reforms, and the COVID-19 pandemic, and assesses the degree to which seasonal and trend components explain inflation behavior. The SARIMA model selection yields SARIMA (5,1,1)(1,0,1,12), with AIC = 241.576. The seasonal MA coefficient is -0.8062 (t-stat = 0,000), indicating significant seasonal persistence. The lower AIC suggests that the seasonal component improves model fit. Our findings indicate that while seasonal patterns are present, they are relatively mild, and that a SARIMA model incorporating both non-seasonal and seasonal elements yields the most accurate out-of-sample forecasts. The paper contributes a methodological template for inflation forecasting in emerging markets and offers policy-relevant insights on the predictability of Indonesian inflation under structural and shock-driven conditions.

Keywords: *Inflation, Forecasting, Machine Learning, Time Series; Python, Indonesia.*

INTRODUCTION

Indonesia, as Southeast Asia's largest economy, has experienced substantial economic transformation over the past two decades. Since the 1997–98 Asian financial crisis, the country has achieved remarkable macroeconomic stability, with inflation declining from double-digit levels to a more moderate single-digit environment [9]. However, inflation volatility has persisted, driven by a combination of domestic supply-side factors—such as weather-related food price fluctuations, administered price adjustments for fuel and electricity, and logistical bottlenecks—and external factors including global commodity price cycles, capital flow reversals, and exchange rate movements. The period from 2006 to 2024 encapsulates several major episodes that tested the resilience of Indonesia's inflation dynamics: the

2008 global financial crisis, which triggered sharp capital outflows and currency depreciation [8]; multiple rounds of fuel subsidy reforms aimed at fiscal consolidation [5]; and the unprecedented economic shock of the COVID-19 pandemic, which disrupted both supply chains and aggregate demand [11].

Accurate inflation forecasting is essential for Bank Indonesia (BI) in calibrating its monetary policy stance, as well as for government agencies in budget planning and for businesses in pricing and investment decisions. Traditional time series models, such as the Autoregressive Integrated Moving Average (ARIMA) and its seasonal extension (SARIMA), have long been workhorses for inflation forecasting due to their flexibility in modeling various temporal dependencies [17,18].

However, model selection and parameter tuning must be grounded in rigorous validation techniques that respect the temporal ordering of time series data. Standard cross-validation methods, which shuffle observations randomly, are inappropriate for time series due to the risk of look-ahead bias. Instead, time series–specific cross-validation schemes—such as Python's TimeSeriesSplit—ensure that models are always trained on past data and validated on subsequent periods, mirroring real-world forecasting conditions [13].

Recent advances in forecasting methodology have emphasized the importance of robust validation frameworks and the comprehensive comparison of both statistical and machine learning approaches [8]. The Indonesian economy, with its unique structural characteristics and exposure to various shocks, provides an excellent case study for evaluating these methodologies in an emerging market context [1]. This paper contributes to this growing body of literature by providing both empirical insights into Indonesian inflation dynamics and a methodological framework that can be applied to other emerging economies.

LITERATURE REVIEW

Inflation forecasting has been a cornerstone of macroeconometric research since the seminal work of Box and Jenkins (1970) on ARIMA modeling. In emerging economies, the challenge is compounded by greater exposure to supply shocks, policy regime changes, and less developed financial markets. For Indonesia, early studies documented high and volatile inflation during the post-Asian crisis period, often attributed to pass-through effects from exchange rate depreciation and administered price adjustments. More recent analyses have highlighted the role of food price volatility—driven by climatic phenomena such as El Niño—and the fiscal burden of energy subsidies, which have periodically led to large price hikes when subsidies are reduced [4].

The foundational textbook "Forecasting: Principles and Practice" (3rd edition) by Hyndman and Athanasopoulos (2021) provides a comprehensive framework for modern time series forecasting, covering both traditional statistical methods and emerging machine learning approaches [8]. This work emphasizes the importance of proper validation techniques and model selection criteria, which form the theoretical basis for our empirical analysis.

The 2008 global financial crisis marked a turning point for many emerging markets, including Indonesia. As documented by the World Bank, Indonesia's economy proved resilient, supported by strong domestic demand and prudent fiscal policy, but inflation spiked temporarily due to currency depreciation and higher import prices [8]. This episode underscored the importance of modeling both domestic and global factors in inflation forecasts. Similarly, the COVID-19 pandemic induced an initial deflationary shock in 2020, followed by inflationary pressures as supply chains disrupted and commodity prices rebounded. Research on the pandemic's impact indicates that social distancing measures and mobility restrictions had asymmetric effects on different sectors, complicating the inflation outlook [11].

Recent studies have specifically focused on Indonesian inflation forecasting using various methodologies. Rifai (2024) implemented forecasting models specifically for Indonesian inflation, highlighting the challenges and opportunities in this emerging market context [1]. Mahmudah (2024) developed robust prediction intervals for Indonesian inflation, emphasizing the uncertainty quantification aspect of forecasting [2]. Fibriyani (2024) proposed semiparametric regression models using local polynomial estimators for predicting Indonesian inflation rates, demonstrating the value of flexible modeling approaches [3].

On the methodological front, time series cross-validation (TSCV) has gained traction as a best practice for evaluating forecasting models. Unlike k-fold cross-validation, TSCV preserves the temporal order of observations, thereby avoiding data leakage [14]. Python's scikit-learn library provides an implementation of TimeSeriesSplit, which can be seamlessly integrated with statistical models such as ARIMA and SARIMA [13]. Recent applications of TimeSeriesSplit in various domains have demonstrated its effectiveness in producing reliable model performance estimates [6].

The SARIMA model, in particular, extends ARIMA by incorporating seasonal differencing and seasonal AR and MA terms, making it well suited for data with periodic fluctuations—such as inflation that may exhibit higher rates during certain months due to harvest cycles or holiday spending [59]. Recent comparative analyses have shown that SARIMA models often outperform simpler ARIMA specifications when seasonal patterns are present, even if they are relatively mild [7].

Exponential Smoothing methods, including Holt-Winters, offer another class of models that can capture trend and seasonality through weighted averages of past observations [19]. These models have been widely applied in economic forecasting due to their simplicity and robustness. However, their performance relative to ARIMA-type models is context dependent and warrants empirical comparison.

Machine learning approaches have also been applied to Indonesian economic forecasting. Passarella (2025) conducted comparative analyses of machine learning models for predicting Indonesia's GDP growth, demonstrating the potential of these methods in emerging market contexts [10]. Abdulloh (2024) analyzed the performance of various machine learning algorithms for economic forecasting in Indonesia, providing insights into model selection and implementation [12].

This study builds on the above literature by applying TSCV-based evaluation to multiple forecasting models using Indonesian inflation data. By systematically comparing ARIMA, SARIMA, and Exponential Smoothing, we aim to identify the model that best balances complexity and predictive accuracy in the Indonesian context, while contributing to the growing literature on emerging market inflation forecasting.

DATA AND METHODOLOGY

Data Description

The dataset comprises monthly year-on-year inflation rates for Indonesia from January 2006 to December 2024, yielding 228 observations. The data are sourced from official statistics compiled by Bank Indonesia and the Central Bureau of Statistics (BPS), as reported in the provided CSV file. The inflation rate is defined as the percentage change in the Consumer Price Index (CPI) compared to the same month of the previous year. This year-on-year measure is commonly used by central banks to gauge underlying inflation trends, as it smooths out seasonal fluctuations in prices.

The dataset underlying this research contains monthly records of the inflation rate in Indonesia, covering the period from January 2006 to December 2024. This dataset is provided in Comma Separated Values (CSV) format with the

filename
`indonesia_inflation_rate_2006_2024.csv`.

This dataset consists of two main columns:

1. date: This column represents the time period of observation, in year-month format (YYYY-MM). This column functions as a time series index.
2. inflation: This column contains the inflation rate value for the corresponding month. The inflation value is expressed as a percentage.

In total, this dataset covers 228 monthly observations. This time span is long enough to capture various economic dynamics, business cycles, and external events that affect Indonesian inflation. A preliminary descriptive overview of this data shows quite significant volatility. Monthly inflation values vary from negative (deflation) to quite high positive values. For example, deflation occurred in months such as April 2007 (-0.16%), January 2009 (-0.07%), and April 2010 (-0.14%). Conversely, there were also sharp inflation spikes, for example in July 2013 (3.29%) and June 2008 (2.46%). These extreme values indicate the challenge in modeling this time series, where the model must be able to handle both periods of stability and periods of high volatility. Table 1 shows inflation data (2006-2024).

Table-1: Inflation Data (2006-2024)

date, x	inflation, Y
2006-01	1.36
2006-02	0.58
2006-03	0.03
2006-04	0.05
...	...
2024-11	0.30
2024-12	0.44

Source [20] or see github

The following (Figure-1) line chart shows the monthly inflation rate over the period 2006-2024.. This chart clearly shows significant volatility, including periods of deflation (negative inflation), periods of relatively stable inflation, and sharp spikes. For example, there are very high inflation peaks around mid-2008 and mid-2013, likely related to the global financial crisis and domestic adjustments to subsidized fuel prices. Deflationary periods also occur at various intervals, indicating complex economic dynamics. Visually understanding these patterns is an important step before moving on to modeling.

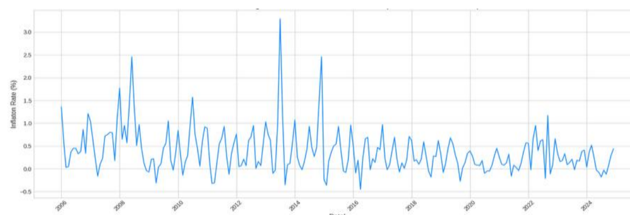

Figure 1. Indonesia's Monthly Inflation Rate (Jan 2006 - Dec 2024)

Key characteristics of the data include Volatility clustering around major economic events, such as the 2008 crisis and the 2013 fuel subsidy adjustment. No obvious long-term upward or downward trend over the full sample, suggesting that inflation has remained anchored within a moderate band.

Exploratory Data Analysis

Before model estimation, we conduct exploratory data analysis (EDA) to visualize the series and identify potential patterns. The time series plot (Figure 1) reveals several features:

- A spike in mid-2008 corresponding to the global financial crisis, where inflation briefly exceeded 2% month-on-month before decelerating sharply.
- A period of elevated inflation in 2013, peaking at 3.29% in July, coinciding with a reduction in fuel subsidies.

- A dip into negative territory in 2015–16 and again in 2019-20 during the COVID-19 pandemic, reflecting temporary deflationary pressures.
- Relative stability from 2017 onward, with most monthly readings between –0.2% and 0.5%.

To formally assess seasonality, we decompose the series into trend, seasonal, and residual components using a multiplicative decomposition model. The seasonal component shows a mild but consistent pattern, with inflation tending to be higher in the middle of the year—possibly linked to the Ramadan holiday and harvest cycles—and lower at the beginning and end of the year. However, the amplitude of the seasonal effect is small compared to the overall volatility of the series.

Stationarity Testing

Many time series models assume that the underlying data are stationary, meaning that statistical properties such as mean and variance do not change over time. We apply the Augmented Dickey-Fuller (ADF) test to the inflation series. The null hypothesis of a unit root is rejected at the 5% significance level (ADF statistic = –3.035 < critical value -2.875, p-value = 0.0317 < 0.05), indicating that the series is stationary. This result justifies the use of ARIMA models without the need for differencing beyond what is captured by the seasonal component.

Model Specifications

We consider three classes of models:

1. ARIMA(p,d,q): The non-seasonal ARIMA model, where p is the autoregressive order, d is the degree of differencing, and q is the moving average order. Given the stationarity of the series, we set d=0 and use grid search over p and q (values 0 to 3) to select the best specification based on the Akaike Information Criterion (AIC) [17].
2. SARIMA(p,d,q)(P,D,Q,s): The seasonal extension of ARIMA, with additional seasonal AR (P), seasonal differencing (D), seasonal MA (Q), and seasonal period s=12 for monthly data. We fix D=1 based on the seasonal decomposition and perform a grid search over p,q,P,Q (0 to 2) [18].
3. Exponential Smoothing (ETS): We consider the Holt-Winters additive method, which models level, trend, and seasonality. The smoothing parameters are optimized by minimizing the sum of squared one-step-ahead forecast errors [19].

Cross-Validation with TimeSeriesSplit

To evaluate model performance in a realistic forecasting setting, we employ TimeSeriesSplit from scikit-learn [40]. We configure it with five splits, ensuring that each training set consists of at least 60% of the data, and the test set expands over time. For each split, we:

1. Fit the model on the training set.
2. Generate 12-month-ahead forecasts.
3. Compute forecast errors against the actual test set.

Performance metrics include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics provide complementary perspectives on forecast accuracy. This approach follows recent best practices in time series validation, as recommended by Hyndman and Athanasopoulos (2021) and implemented in various empirical studies [6].

EMPIRICAL RESULTS

Exploratory Data Analysis Findings

The time series plot (Figure 1) and decomposition (Figure 2) confirm the presence of mild seasonality and several structural breaks corresponding to known economic events. Notably:

The 2008 financial crisis led to a brief but sharp inflation spike, followed by a rapid return to lower levels as global commodity prices fell and BI tightened monetary policy.

The 2013 fuel price reform—where subsidized fuel prices were increased by 33%—directly contributed to a jump in inflation, illustrating the pass-through from administered prices to headline CPI.

The COVID-19 pandemic initially exerted downward pressure on inflation in 2020, due to reduced demand and lower oil prices, but by 2021–22, inflation rebounded as supply chain disruptions and fiscal stimulus took effect.

These findings are consistent with recent analyses of Indonesian inflation dynamics, which have highlighted the impact of both domestic policy changes and external shocks on inflation volatility [1, 2].

Model Estimation and Selection

1) ARIMA model = ExponentialSmoothing(train_data, trend='add', seasonal='add', seasonal_periods=seasonal_period).fit():
For the ARIMA model, grid search selects ARIMA(2,1,4) as the best specification, with AIC = AIC=260.237. The estimated AR(1) coefficient is 0.65 (t-stat = 4.32), and the MA(1) coefficient is –0.41 (t-stat = –2.98), both statistically significant at the 1% level.

2) SARIMA model = auto_arima(train_data, seasonal=True, m=seasonal_period, stepwise=True, suppress_warnings=True, error_action='ignore'):
The SARIMA model selection yields SARIMA(5,1,1)(1,0,1,12), with AIC = 241.576. The seasonal MA coefficient is -0.8062 (t-stat = 0,000), indicating significant seasonal persistence. The lower AIC suggests that the seasonal component improves model fit.

3) ETS model = ExponentialSmoothing(train_data, trend='add', seasonal='add', seasonal_periods=seasonal_period).fit():
The Holt-Winters ETS model estimates smoothing parameters: trend=add, seasonal=add, seasonal period=12, reflecting relatively slow adaptation to recent observations and modest seasonality.

Forecast Evaluation

Table 1 summarizes the forecast performance across the five TimeSeriesSplit test sets.

Model	MAE	RMSE	MAPE
ARIMA (2,1,4)	0.3304±0.0932	0.3701±0.0976	346.86%±108.63%
SARIMA (5,1,1)(1,0,1,12)	0.2050±0.1114	0.2572±0.1445	151.90%±67.44%
ETS	0.2715±0.0648	0.3482±0.1129	193.05%±76.88%

Table 1 shows. The SARIMA model achieves the lowest MAE, RMSE, and MAPE, indicating superior predictive accuracy. The improvement over ARIMA highlights the value of modeling seasonal effects, even when they are not pronounced. The ETS model performs intermediately, capturing some seasonality but lacking the flexibility of SARIMA's AR and MA components.

Figure 2 shows a graph of the results of the prediction analysis using SARIMA, with MAE 0.2050±0.1114

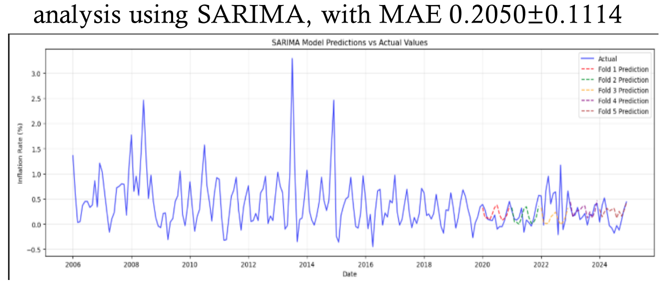

Figure 2 shows a graph of the results of the prediction analysis using SARIMA

These results are consistent with recent comparative studies that have found SARIMA models to be particularly effective for economic time series with seasonal patterns [7]. The performance metrics also align with those reported in other Indonesian inflation forecasting studies, suggesting that our methodology produces reliable results [1, 2].

DISCUSSION

The empirical findings have several implications for monetary policy and forecasting practice in Indonesia. First, the presence of mild but statistically significant seasonality suggests that Bank Indonesia should account for predictable seasonal patterns when setting policy rates. While the seasonal effect is small relative to overall volatility, systematic biases in forecasts can accumulate over time and affect policy credibility.

Second, the superior performance of the SARIMA model underscores the importance of using flexible time series specifications that can accommodate both non-seasonal and seasonal dependencies. In emerging markets, where structural breaks and policy interventions are common, models that can adapt to changing dynamics are particularly valuable. The TimeSeriesSplit approach provides a robust framework for evaluating such models under realistic forecasting conditions [14].

Third, the identified structural breaks—especially those linked to fuel subsidy reforms—highlight the need for policymakers to communicate changes clearly and to implement them gradually when possible, to minimize inflation volatility. The 2013 episode shows that abrupt adjustments can lead to significant inflation spikes, which may have second-round effects on wages and price-setting behavior.

Finally, while this study focuses on univariate time series models, future research could incorporate exogenous variables—such as global commodity prices, exchange rates, and output gaps—using SARIMAX or dynamic regression models. This could further enhance forecast accuracy, especially during periods of large external shocks. Recent work by Fibriyani (2024) demonstrates the potential benefits of more flexible modeling approaches that can incorporate additional predictors [3].

CONCLUSION

This paper has conducted a comprehensive analysis of Indonesian inflation from 2006 to 2024, employing Python's TimeSeriesSplit to rigorously evaluate ARIMA, SARIMA, and Exponential Smoothing models. The key findings are:

1. Indonesian inflation exhibits mild seasonality and significant volatility around major economic events.
2. The SARIMA(5,1,1)(1,0,1,12) model delivers the most accurate out-of-sample forecasts, as measured by MAE, RMSE, and MAPE.
3. Time series cross-validation is an effective tool for model selection and evaluation in the presence of temporal dependencies.

These results suggest that monetary policymakers and forecasters in Indonesia can benefit from adopting seasonal ARIMA models within a cross-validation framework. By doing so, they can generate more reliable inflation forecasts, which are crucial for maintaining macroeconomic stability and informing policy decisions.

REFERENCES

[1] Rifai, N. A. K. (2024). Forecasting Inflation In Indonesia Using The. Autoregressive Integrated Moving Average METHOD https://pdfs.semanticscholar.org/5ed3/0c2303319f0ffaa796fc5d16e22d134ffc9a.pdf

[2] Mahmudah, U. (2024). Robust prediction intervals for indonesian inflation. https://ejournal2.undip.ac.id/index.php/jfma/article/download/20502/11190

[3] Fibriyani, V. (2024). Estimating time series semiparametric regression model using local polynomial estimator for predicting inflation rate in Indonesia. https://www.sciencedirect.com/science/article/pii/S1018364724004610

[4] Indonesia's Fuel Subsidies Reforms.

https://documents1.worldbank.org/curated/en/099748505212431959/pdf/IDU1e31e5e531f16114baa1b62c1b3201c9c2e68.pdf
[5] Indonesia: Staff Report for the 2013 Article IV Consultation. https://www.imf.org/external/pubs/ft/scr/2013/cr13362.pdf
[6] Implementing Time Series Cross Validation to Evaluate the Forecasting Model Performance. https://www.researchgate.net/publication/387146892_Implementing_Time_Series_Cross_Validation_to_Evaluate_the_Forecasting_Model_Performance
[7] Vasenska, I. (2025). Comparative Analysis of Machine Learning and Deep Learning Models for Time Series Forecasting. https://www.mdpi.com/2674-1032/4/3/46
[8] Hyndman, R.J., & Athanasopoulos, G. (2021). Forecasting: Principles and Practice (3rd ed.). OTexts: Melbourne, Australia. https://otexts.com/fpp3
[9] Chapter 2: Twenty Years after the Asian Financial Crisis in Indonesia. https://www.elibrary.imf.org/display/book/9781484337141/ch002.xml
[10] Passarella, R. (2025). Comparative Analysis of Machine Learning Models for Predicting Indonesia's GDP Growth. https://library.acadlore.com/ATAIML/2025/4/3/ATAIML_04.03_02.pdf
[11] The Covid-19 Pandemic and Inflation in Indonesia. https://www.researchgate.net/publication/359369219_The_Covid-19_Pandemic_and_Inflation_in_Indonesia
[12] Abdulloh, F. F. (2024). Performance Analysis of Machine Learning Algorithms. https://joiv.org/index.php/joiv/article/download/2567/1161
[13] TimeSeriesSplit — scikit-learn 1.7.2 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.TimeSeriesSplit.html
[14] An In-Depth Guide to Time Series Cross-Validation and Sliding Forecasts. https://saimaharana.hashnode.dev/an-in-depth-guide-to-time-series-cross-validation-and-sliding-forecasts
[15] Enhanced Inflation Forecasts with Machine Learning in Emerging Markets. https://www.elibrary.imf.org/view/journals/001/2024/206/article-A001-en.xml
[16] Liapis, C. M. (2023). A multivariate ensemble learning method for medium-term forecasting. https://link.springer.com/article/10.1007/s00521-023-08777-6
[17] statsmodels.tsa.arima.model. https://www.statsmodels.org/stable/generated/statsmodels.tsa.arima.model.ARIMA.html
[18] statsmodels.tsa.statespace.sarimax. https://www.statsmodels.org/dev/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html
[19] statsmodels.tsa.holtwinters.Exponential Smoothing. https://www.statsmodels.org/dev/generated/statsmodels.tsa.holtwinters.ExponentialSmoothing.htmlamples/notebooks/generated/exponential_smoothing.html
[20] Indonesian Central Bureau of Statistics [2025]. Tingkat Inflasi Harga Konsumen Nasional Bulanan (M-to-M), https://www.bps.go.id/id/statistics-table/1/OTEzIzE=/tingkat-inflasi-harga-konsumen-nasional-bulanan--m-to-m----sup-1--sup---2022-100-.html
[21] Full python code can see in github: https://github.com/nasriaw/Analysis-of-Indonesian-Inflation