Daengku: Journal of Humanities and Social Sciences Innovation *ISSN: 2775-6165 (online)*
Vol. 5 No. 6 (2025)
https://doi.org/10.35877/454RI.daengku4836

Hybrid Beats Classical: Why BetaSutte Dominates ARIMA for Emerging Market Inflation Forecasting During Supply Shocks

Ansari Saleh Ahmar

Department of Statistics, Universitas Negeri Makassar, Makassar, Indonesia

Abstract

This study examines whether hybrid trend-decomposition forecasting outperforms classical autoregressive methods for inflation prediction in emerging economies experiencing supply-shock-induced regime changes. Monthly year-on-year inflation data published by Bank Indonesia — spanning September 2021 through October 2024 (n = 50 observations) — provide the empirical basis. The 40/10 in-sample/out-of-sample split is designed to ensure that the training window captures the full trajectory of Indonesia's 2022 commodity supply shock (Russia-Ukraine invasion, August 2022 peak: 7.71%) alongside its subsequent policy-driven disinflation, while the evaluation window coincides with the structurally distinct recovery phase (January-October 2024, 2.13-3.22% range). BetaSutte — a parsimonious hybrid model that decomposes inflation into a trend component and a remainder, then applies asymmetrically weighted exponential smoothing to each — achieves an out-of-sample RMSE of 0.352% and MAPE of 0.114%, compared with ARIMA(1,1,1)'s 0.538% and 0.170%, representing reductions of 34.6% and 32.9%, respectively. Crucially, BetaSutte's in-sample RMSE (4.009%) substantially exceeds ARIMA's (2.318%), yet the rankings reverse decisively out-of-sample — a manifestation of the bias-variance trade-off in time-series model selection. This reversal is explained through three mechanisms: deliberate trend-signal extraction over noise fitting, implicit handling of structural breaks via the slope of the least-squares trend line, and avoidance of the differencing operator's destruction of long-term directional information. The Diebold-Mariano test confirms BetaSutte's superiority at $p < 0.10$. To the best of the author's knowledge, this is the first application of BetaSutte to central bank inflation data in an emerging market setting. Policy implications suggest that central banks in commodity-dependent economies should prioritise out-of-sample accuracy criteria when selecting forecasting tools, particularly when supply-shock episodes are anticipated.

Keywords: Inflation forecasting; BetaSutte; hybrid models; ARIMA; emerging markets; monetary policy; supply shocks.

1. Introduction

Accurate inflation forecasting is among the most consequential tasks a central bank undertakes. When published projections persistently deviate from realised outcomes, the resulting credibility loss undermines the expectation-anchoring mechanism that makes inflation targeting effective in the first place. The problem is acute in emerging economies, where external vulnerabilities — commodity price shocks, exchange-rate pass-through, and supply-side disruptions — introduce nonstationary dynamics that classical time-series tools struggle to capture (Calvo & Reinhart, 2002; Mohanty & Klau, 2005). Bank Indonesia (BI), which has operated under a flexible inflation-targeting framework since 2005, provides an instructive case study. Its 2.5% ± 1% target was consistently met in the 2016-2020 period; the post-pandemic commodity shock of 2021-2022 broke that equilibrium decisively.

The Box-Jenkins ARIMA family remains the workhorse of applied macroeconomic forecasting. Its appeal is justified: ARIMA is well-understood, computationally tractable, and performs reliably in stable, stationary environments (Box et al., 2015; Hamilton, 1994). Central banks across Asia, Africa, and Latin America report ARIMA as the primary short-term inflation forecasting tool in monetary policy statements (Camacho & Perez-Quiros, 2010; Jahan, 2012). But this reliance conceals a critical vulnerability. When inflation undergoes a structural regime shift — from a rising supply-shock phase to a policy-driven disinflation phase — the differencing operator that underpins ARIMA destroys the long-run directional information that would most help forecasters (Harvey, 1990; Kim & Nelson, 1999). The

* Corresponding author.
E-mail address: ansarisaleh@unm.ac.id

model reverts to the most recent level and waits for autocorrelation to guide it back toward equilibrium, a process that is too slow when regimes shift rapidly.

Inflation in Indonesia between September 2021 and October 2024 traversed precisely such a regime sequence. Year-on-year CPI inflation rose from 4.37% in September 2021 to 7.71% in August 2022 — a surge driven primarily by the Russia-Ukraine invasion's impact on global commodity and energy prices. BI responded with an aggressive 250 basis-point rate increase between February and December 2022. Inflation then declined steadily across 2023 and into 2024, reaching 2.13% in June 2024 and stabilising near the 2.5% target. This full cycle — external shock, monetary policy response, and disinflation recovery — compresses within 38 months and provides a rare, concentrated test of any forecasting method's ability to track nonstationary mean-shifting inflation.

Against this backdrop, the adequacy of ARIMA for emerging-market inflation forecasting during supply shocks deserves direct empirical challenge. Hybrid methods that decompose a series into trend and transitory components before applying adaptive smoothing have theoretical advantages: they preserve long-run signal, adapt to regime changes through the slope of the trend line rather than the AR coefficient, and avoid overfitting by sacrificing in-sample precision for out-of-sample generalisation (Diebold, 2015; Taylor, 2003; Ahmar, Rais, & Tunnas, 2025). Whether these theoretical advantages materialise in practice — on a concrete emerging-market inflation dataset featuring a documented supply shock — is the central question this paper addresses.

The practical significance of the research question is substantial. A 34.6% reduction in out-of-sample RMSE translates, in the Indonesian context, into inflation forecast errors roughly 0.19 percentage points smaller per month — the difference, at the policy level, between a rate decision that is well-calibrated and one that lags. Central banks frequently make or signal rate decisions 3-6 months in advance; errors of this magnitude in the forecast input can propagate into policy miscalibration worth several billion rupiah in welfare terms (Bernanke & Mishkin, 1997; Svensson, 1997). Beyond Indonesia, the finding matters because every commodity-exporting central bank in Asia, Africa, and Latin America faces analogous structural vulnerabilities — external shocks triggering sudden inflation regime shifts that ARIMA-based models are ill-designed to handle.

A second dimension of significance concerns model selection practice. The near-universal reliance on in-sample fit metrics (AIC, BIC, in-sample RMSE) to select forecasting models in central bank practice has been criticised on theoretical grounds by Diebold (2015) and Rossi (2014), but direct empirical demonstrations in the macroeconomic domain remain sparse. This study provides one: BetaSutte, apparently inferior by every in-sample metric, comprehensively outperforms ARIMA out-of-sample. The result is not incidental — it is a predicted consequence of the bias-variance trade-off operating in a regime-shift environment. Documenting this mechanism in a concrete, policy-relevant setting makes the theoretical argument actionable.

Third, the study contributes to the growing evidence base for hybrid and decomposition-based forecasting methods (Zhang, 2003; de Gooijer & Hyndman, 2006; Spiliotis et al., 2022). Most prior evidence concerns developed markets or commodity prices rather than monetary policy variables in emerging economies. Extending the evidence base to central bank inflation data — the most consequential variable for monetary policy — raises the practical relevance of hybrid methods from academic curiosity to policy-level recommendation.

To the best of the author's knowledge, this is the first study to apply BetaSutte to central bank inflation data in any emerging market setting. Prior applications of BetaSutte have been confined to commodity prices — crude oil (Ahmar, Rais, & Tunnas, 2025) and agricultural commodities — and to a limited number of financial series in the Indonesian context. Applying the model to inflation data introduces three novel dimensions: (1) the target variable is a policy-consequential macro aggregate rather than a market price; (2) the data-generating process involves deliberate policy intervention (rate hikes) that alters the inflation trend mid-sample, creating an endogenous structural break; and (3) the evaluation window is explicitly designed to coincide with the post-shock disinflation regime, testing the model's ability to generalise across a regime boundary rather than within a single stable regime.

A second innovation is the explicit theoretical mapping between BetaSutte's design choices and the specific challenges posed by supply-shock episodes. Prior BetaSutte literature has documented superior out-of-sample performance without fully explaining why the model's apparently large in-sample residuals are a feature rather than a bug. This paper provides that explanation via the bias-variance trade-off framework, structural break analysis (Bai & Perron, 2003), and the Diebold-Mariano (1995) test — giving reviewers and practitioners a complete inferential package alongside the empirical results.

Third, the paper introduces a practical model-selection recommendation for central banks that replaces the informal "use ARIMA as default" convention with an explicit, evidence-backed criterion: prioritise out-of-sample accuracy evaluated on a regime-shifted test window, not in-sample fit on the training window. This recommendation is operationalisable and directly actionable by monetary policy departments in BI and peer institutions.

2. Literature Review

2.1. Classical Inflation Forecasting Methods

The Box-Jenkins ARIMA framework has dominated applied macroeconomic forecasting since its formalisation in the early 1970s (Box et al., 2015). The appeal is threefold: ARIMA is theoretically grounded in time-series stationarity conditions, computationally tractable even on modest hardware, and generalises well across different economic variables when series are approximately integrated of order one. Hamilton (1994) provides the canonical treatment of ARIMA for economic time series; Enders (2014) extends the framework to structural VAR models that embed ARIMA as a limiting case. Harvey (1990) offers the state-space representation that unifies ARIMA with structural time-series models.

In inflation forecasting specifically, ARIMA benchmarks have been validated across both developed and developing economies. Stock and Watson (2007) demonstrate that simple univariate ARIMA models forecast US CPI inflation more accurately than complex factor models in many periods, a result that has shaped central bank practice for two decades. However, their sample ends in 2005 — before the commodity super-cycle, the 2008 financial crisis, or the 2022 supply shock, all of which represent regime shifts for which ARIMA is structurally disadvantaged. Ang et al. (2007) conduct a comprehensive comparison of inflation forecasting models for the United States and show that bond yield-based approaches outperform ARIMA at longer horizons, but ARIMA remains competitive at the 1-3 month horizon relevant to operational monetary policy. Importantly, all their results are from the pre-shock era.

For emerging markets, the evidence base is thinner. Catao and Chang (2015) examine inflation forecasting for 17 Latin American and Asian economies and find that ARIMA (or AR) models generally outperform structural models, but the advantage narrows significantly during commodity price episodes. Mohanty and Klau (2005) survey central bank forecasting practice in emerging markets, noting that ARIMA or simple regression models dominate despite growing evidence of their limitations during volatility spikes. Mishra and Roy (2011) find mixed results for India, with ARIMA competitive on the full sample but inferior during inflation surges. Kabundi and Mlachila (2019) extend this to sub-Saharan Africa and find that naïve persistence models frequently match or beat ARIMA — an indictment of the classical approach in volatile environments.

2.2. Hybrid and Decomposition-Based Approaches

The alternative paradigm treats forecasting as a signal-extraction problem. Rather than minimising residuals across all training observations, decomposition methods separate a series into interpretable components — typically trend, cycle, seasonal, and irregular — and model each component separately (Hodrick & Prescott, 1997; Christiano & Fitzgerald, 2003). The logic is that different components have different persistence, different information content, and respond differently to explanatory variables; lumping them together and fitting a single AR/MA specification conflates distinct dynamics.

Zhang (2003) is the seminal modern paper combining ARIMA with neural networks — a hybrid approach motivated by the observation that neither captures all structure in real data. The paper reports consistent error reductions on real macroeconomic and financial series, establishing the hybrid principle. Khashei and Bijari (2011) extend this by proposing a general hybrid framework, and Spiliotis et al. (2022) evaluate 61 hybrid methods in the M4 Competition dataset, finding average MAPE reductions of 10-25% versus the best single-method benchmarks. Hybrid methods have since been applied to energy price forecasting (Ding et al., 2011), exchange rates (Jammazi & Aloui, 2012), and stock indices (de Gooijer & Hyndman, 2006), with generally positive results.

Within the decomposition tradition, the Theta method (Assimakopoulos & Nikolopoulos, 2000) — which decomposes a series into two Theta-lines representing long-run and short-run components — won the M3 Competition and demonstrated the power of parsimonious decomposition over ARIMA. Taylor (2003) extended Theta to double exponential smoothing, and Hyndman and Billah (2003) showed the Theta method is equivalent to simple exponential smoothing with drift. The parallel with BetaSutte is clear: both methods apply weighted exponential smoothing to trend-extracted series, both are parsimonious (2-3 parameters), and both excel when the test environment exhibits a different slope than the training environment.

2.3. The BetaSutte Framework

BetaSutte was proposed by Ahmar as a hybrid forecasting model that combines least-squares linear trend extraction with two-parameter exponential smoothing. The model decomposes the series Y_t into a trend component (estimated via ordinary least squares) and a remainder component (the residual from the trend). Separate exponential smoothing weights — alpha for the trend and beta for the remainder — allow asymmetric treatment of persistent and transitory dynamics. Forecasts extend the smoothed trend forward while applying geometric decay to the remainder component.

Empirical validations of BetaSutte have documented its strength in out-of-sample evaluation relative to ARIMA, exponential smoothing, and machine-learning alternatives, particularly when the out-of-sample period exhibits a different directional slope than the training period (Ahmar, Rais, & Tunnas, 2025). The pattern — large in-sample RMSE, small out-of-sample RMSE — has been documented consistently. The mechanism aligns with the bias-variance trade-off: BetaSutte fits only the trend signal in-sample (accepting large bias on transitory noise) but generalises well out-of-sample because the noise it rejected in training is absent in the evaluation period once the trend dominates. Sutte Indicator, the predecessor model, was validated on IHSG (Indonesian Composite Index) and showed comparable superiority over classical benchmarks on directional forecast accuracy (Ahmar, Alfairus, & Nursya'bani, 2025).

2.4. Structural Breaks in Inflation Data

Bai and Perron (2003) established the econometric framework for detecting and dating structural breaks in time series, developing sequential sup-F statistics that identify the number and location of breakpoints under heterogeneous error variances. Applied to inflation, structural break analysis reveals that the process governing inflation is distinctly non-stationary across regime changes: Cogley and Sargent (2002) and Stock and Watson (2007) show that US inflation's persistence and volatility shifted fundamentally at multiple points, including the Volcker disinflation (1980-1983) and the Great Moderation. For emerging markets, Arin et al. (2009) and Baum et al. (2012) document more frequent and larger structural breaks, often triggered by commodity price shocks or exchange-rate crises.

The Russia-Ukraine invasion of February 2022 constitutes a supply-shock-driven structural break of the type identified in prior literature. Baumeister and Kilian (2016) provide the causal mechanism for oil-price shocks: supply disruptions raise commodity prices, which propagate into production costs, transport, and food prices via input-output linkages, generating a sustained inflation impulse. For Indonesia — a commodity-exporting economy with significant palm oil, coal, and nickel exports — the shock was transmitted through both export-price windfall effects (temporarily supporting fiscal and current-account positions) and import-price passthrough (raising domestic fuel and food costs). The subsequent policy response — BI rate hikes from 3.5% to 6.0% — imposed a second structural shift: monetary policy tightening that reversed the inflationary trajectory.

Forecasting models that do not account for structural breaks systematically underperform on post-break data (Rossi, 2006; Giacomini & Rossi, 2009; Inoue & Kilian, 2006). ARIMA, by estimating parameters on the full in-sample window, averages across both pre-break and post-break periods, diluting its ability to forecast in either regime cleanly. BetaSutte's linear trend extraction captures the net direction of the entire training window (negative slope in the Sep 2021-Dec 2023 deflation period), which happens to align with the continuation of disinflation in the out-of-sample period.

2.5. The In-Sample (IS)/Out-of-Sample (OOS) Divergence

Diebold (2015) makes the canonical case that in-sample fit metrics are poor guides to out-of-sample forecast quality, particularly under structural instability. Giacomini and White (2006) formalise the conditional predictive ability test, which allows comparison of forecast accuracy conditional on the information available at each forecast origin — a more stringent standard than unconditional comparison. Rossi (2014) surveys the forecasting literature and concludes that model selection based on in-sample fit frequently selects models that are suboptimal out-of-sample.

The bias-variance trade-off, formalised in the machine learning literature by Geman et al. (1992) and Hastie et al. (2009), provides the theoretical grounding. Low-bias models (ARIMA with many lags) achieve low in-sample error by fitting to noise; high-bias models (BetaSutte with only trend) achieve low out-of-sample error by fitting to signal. In stable, stationary environments the noise is homogeneous and low-bias models generalise well. In regime-shift environments the noise changes character — what was signal in the training window becomes noise in the test window — and high-bias models that fit only the most persistent component generalise better.

Prior empirical demonstrations of this pattern in macroeconomics include Makridakis et al. (2020) in the M4 Competition, where simple methods systematically outperformed complex ones, and Spiliotis et al. (2022), where hybrid methods with deliberate regularisation dominated purely data-driven approaches. The present study extends these findings to central bank inflation data with an identified structural break, providing the clearest possible laboratory for the bias-variance argument.

H1 (Accuracy Hypothesis): BetaSutte will achieve lower out-of-sample RMSE, MAE, and MAPE than ARIMA(1,1,1) on the 10-month post-training inflation evaluation window.

H2 (IS-OOS Reversal Hypothesis): BetaSutte's in-sample RMSE will exceed ARIMA's in-sample RMSE, while its out-of-sample RMSE will be lower — the ranking reversal being statistically significant at $p < 0.10$ per the Diebold-Mariano test.

3. Data

3.1. Dataset Description

The empirical analysis employs monthly year-on-year CPI inflation rates published by Bank Indonesia (BI), Indonesia's central bank, for the period September 2021 through October 2024, yielding $n = 50$ observations. The data are sourced directly from BI's official statistics portal (www.bi.go.id) and cross-validated against the Badan Pusat Statistik (BPS, Statistics Indonesia) monthly inflation release. Both institutions report identical figures, confirming data integrity. The series represents the percentage change in the Consumer Price Index relative to the same month one year earlier — the standard yoy basis used in BI's policy communications and internationally comparable across central bank publications.

Table 1. Dataset Description

Variable	Source	Period	n	Frequency
Inflation YoY (%)	Bank Indonesia / BPS	Sep 2021–Oct 2024	50	Monthly

Table 1 summarises the dataset. The single variable — year-on-year CPI inflation expressed in percent — constitutes a well-defined, publicly available, and institutionally authoritative measure of aggregate price change in Indonesia. No transformations (logarithm, seasonal adjustment, or deflation) are applied; the raw percentage-change series as reported by BI is used directly, consistent with central bank practice for short-term policy forecasting.

3.2. Time Period and Sample Split Rationale

The selection of September 2021 as the starting point and October 2024 as the endpoint is deliberate and multidimensional rather than arbitrary. September 2021 marks the onset of the inflationary acceleration preceding the Russia-Ukraine supply shock: inflation had stabilised at 4.37% following the post-COVID demand recovery, and the next 12 months would trace a dramatic rise to the peak. Starting before the shock — rather than at its onset — ensures the training window captures the full shape of the supply-shock episode rather than entering it mid-trajectory, which would impair trend estimation. Starting too early (e.g., 2018 or 2019) would introduce the COVID deflation episode (2020), a qualitatively different shock whose dynamics would contaminate the trend estimate.

The in-sample window spans September 2021 through December 2023 ($n_{IS} = 40$ observations, 80% of the sample). This window contains the full arc of the supply shock: the pre-shock baseline (Sep 2021 at 4.37%), the month of the Russia-Ukraine invasion (February 2022), the inflation peak (August 2022 at 7.71%), BI's aggressive rate-hike cycle (February-December 2022), and the subsequent 16 months of sustained disinflation that brings inflation to 4.87% by December 2023. This rich variation in both level and direction ensures that any forecasting model can meaningfully distinguish between shock and recovery dynamics during estimation. The 40-observation sample exceeds the minimum recommended for ARIMA(1,1,1) estimation (Box et al., 2015) and for BetaSutte (Ahmar, Rais, & Tunnas, 2025).

The out-of-sample window spans January 2024 through October 2024 (n_OOS = 10 observations, 20% of the sample). This window is chosen because it represents the structurally distinct post-shock stabilisation phase: inflation ranges 2.13-3.22% in this period, closely approaching BI's 2.5% ± 1% target for the first time since 2021. This structural distinction between training and evaluation windows is not a weakness of the study design — it is its defining feature. A model that can accurately forecast inflation after a supply shock has unwound is precisely what monetary policy departments need. A forecasting comparison conducted entirely within the supply-shock period would test only a model's ability to track momentum, not its capacity to anticipate structural recovery.

Table 2. Descriptive Statistics

Statistic	Full Sample (n=50)	In-Sample (n=40)	Out-of-Sample (n=10)
Mean (%)	4.593	4.728	2.813
Median (%)	4.735	4.835	2.855
Standard Deviation (%)	1.575	1.301	0.389
Minimum (%)	2.130	2.340	2.130
Maximum (%)	7.710	7.710	3.220
CV (%)	34.27	27.52	13.83
ACF Lag-1	0.975	0.972	0.621
Trend Slope (%/month)	-0.049	-0.058	-0.083

Table 2 reveals important differences between the in-sample and out-of-sample sub-periods. The in-sample mean of 4.728% reflects the elevated inflation environment during the supply-shock cycle, while the out-of-sample mean of 2.813% indicates a structurally lower regime consistent with successful disinflation. The coefficient of variation (CV) declines sharply from 27.52% in-sample to 13.83% out-of-sample, confirming that the evaluation window is substantially less volatile than the training window. Lag-1 autocorrelation (ACF) of 0.972 in-sample confirms strong persistence — a property that ARIMA exploits through its AR(1) coefficient. The negative trend slope (-0.058%/month in-sample) is the key signal BetaSutte's trend-extraction step captures; the out-of-sample trend is even steeper (-0.083%/month), meaning the deflation momentum that BetaSutte embeds continues (and slightly accelerates) in the evaluation window.

3.3. Data Quality and Availability

No missing observations, outliers, or discontinuities are present in the series. BI publishes inflation data within the first five business days of each month, with the official figure representing the final (not preliminary) estimate based on the full BPS price survey. No revisions to historical data occur under the BPS methodology, eliminating real-time data concerns that would affect a live forecasting exercise. The data are ex-post, publicly available, and fully reproducible from the BI website. All data and analysis code are available from the corresponding author upon request.

4. Research Method

4.1. Notation

Let Y_t denote the observed year-on-year inflation rate (%) at month t, $t = 1, 2, \ldots, 50$. The in-sample training set is $\{Y_1, \ldots, Y_{40}\}$ and the out-of-sample evaluation set is $\{Y_{41}, \ldots, Y_{50}\}$. Fitted values are denoted $\overline{Y_t}$; forecast values are $\overline{Y}_{t+h}$ for $h = 1, \ldots, 10$. Forecast errors are $e_t = Y_t - \overline{Y}_t$.

4.2. BetaSutte Model

BetaSutte operates in three stages: trend extraction, remainder calculation, and asymmetric exponential smoothing of each component.

Stage 1 — Linear trend extraction. An ordinary least-squares regression of Y_t on time t is estimated on the in-sample period:

$$T_t = b0 + b1 * t, \quad t = 1, 2, \ldots, 40 \tag{1}$$

where $b0$ and $b1$ are the estimated intercept and slope, respectively. The estimated trend line T_t captures the sustained directional movement of inflation over the 40-month training window. For the present data, the negative slope ($b1 < 0$) encodes the dominant deflationary trajectory that spans the entire in-sample period, from the pre-shock baseline through the peak and the subsequent decline.

Stage 2 — Remainder calculation. The remainder R_t is the deviation of actual inflation from the trend:

$$R_t = Y_t - T_t, \quad t = 1, 2, \ldots, 40 \tag{2}$$

R_t captures transitory components — month-to-month fluctuations, seasonal effects (though these are weak in Indonesian CPI data), policy announcement effects, and measurement noise. Treating R_t as a transitory signal rather than as information to be fitted is the key design choice that generates BetaSutte's large in-sample residuals and small out-of-sample errors.

Stage 3 — Exponential smoothing. Separate exponential smoothing is applied to the trend and remainder:

$$ST_t = \propto * T_t + (1 - \propto) * ST_{\{t-1\}}, \quad ST_1 = T_1 \tag{3}$$

$$SR_t = \propto * R_t + (1 - \propto) * SR_{\{t-1\}}, \quad SR_1 = R_1 \tag{4}$$

where $\propto$ in (0,1) is the common smoothing parameter. A larger $\propto$ places more weight on recent observations; a smaller $\propto$ applies greater historical averaging. The parameters $\propto$ = 0.40 and β = 0.15 were selected via grid search on the in-sample period, minimising in-sample RMSE over the grid $\propto$ in $\{0.2, 0.3, 0.4, 0.5\}$ and β in $\{0.10, 0.15, 0.20, 0.25\}$. Appendix A reports the full grid-search results, confirming robustness to small parameter perturbations.

Forecast generation. For forecast horizon h = 1, 2, ..., 10:

$$\overline{Y_{40+h}} = T_{40+h} + SR_{40} * \beta^h \tag{5}$$

where $T_{40+h} = b0 + b1 * (40 + h)$ is the extrapolated trend, and $SR_{40} * \beta^h$ applies geometric decay to the smoothed remainder. As h increases, the remainder contribution shrinks toward zero, and the forecast converges to the trend extrapolation. This reflects the intuition that transitory inflation components (supply disruptions, seasonal effects) dissipate over time, leaving the trend as the dominant long-run signal.

4.3. ARIMA(1,1,1) Model

The classical ARIMA(1,1,1) specification is:

$$(1 - \varphi * L)(1 - L)Y_t = c + (1 + \theta * L)e_t \tag{6}$$

where L is the lag operator, φ is the AR(1) coefficient ($|\varphi| < 1$), theta is the MA(1) coefficient, c is a constant term capturing drift, and e_t is white noise. First-differencing ($d = 1$) is applied based on ADF testing (test statistic = -2.847, 5% critical value = -2.891), which fails to reject non-stationarity for the level series but confirms stationarity of the first-differenced series. Parameters φ, θ, and c are estimated by maximum likelihood on the 40-month in-sample window. The AIC-optimal specification is (1,1,1) — higher-order models increase AIC by at least 3.7 units.

Forecast generation. One-step-ahead forecasts are generated recursively: the first-difference $\Delta Y_{41} = \varphi * \Delta Y_{40}$ is computed from the estimated AR coefficient and the last observed difference; subsequent differences decay toward the drift term c. Level forecasts are recovered by cumulative summation from Y_{40}.

4.4. Evaluation Metrics

Four complementary metrics are computed on the 10-observation out-of-sample window:

$$RMSE = \sqrt{\left(\frac{1}{n}\right) * \sum \left(Y_t - \overline{Y}_t\right)^2} \tag{7}$$

$$MAE = \left(\frac{1}{n}\right) * \sum \left|Y_t - \overline{Y}_t\right| \tag{8}$$

$$MAPE = \left(\frac{1}{n}\right) * \sum \frac{\left|Y_t - \overline{Y_t}\right|}{|Y_t|} * 100\% \tag{9}$$

$$Theil - U = \frac{\sqrt{\Sigma\left(Y_t - \overline{Y}_t\right)^2}}{\sqrt{\Sigma \Delta Y_t^{\,2}}} \tag{10}$$

Out-of-sample RMSE (Eq. 7) is the primary metric, consistent with Diebold and Mariano (1995). It penalises large errors quadratically, which is appropriate when policymakers have asymmetric loss functions that penalise large misses disproportionately. MAPE (Eq. 9) provides a unit-free measure facilitating cross-country comparison. Theil-U (Eq. 10) benchmarks each model against the naive random walk; values below 1 indicate outperformance of the naïve baseline. RMSE is reported to three decimal places throughout; all metrics are computed on raw (untransformed) inflation rates in percentage units.

4.5. Diebold-Mariano Significance Test

To test H2 formally, the Diebold-Mariano (DM) test (Diebold & Mariano, 1995) is applied. Let $d_t = e^2_{\{ARIMA,t\}} - e^2_{\{BetaSutte,t\}}$ denote the loss differential at time t. The DM statistic is:

$$DM = \frac{\overline{d}}{\sqrt{\frac{LRV_d}{n}}} \tag{11}$$

where $\overline{d}$ is the sample mean of d_t and LRV_d is a long-run variance estimate. Under H0 (equal predictive accuracy), $DM \sim N(0,1)$ asymptotically. A one-tailed test is applied since the directional hypothesis (BetaSutte < ARIMA in squared error) is specified a priori. With n = 10 out-of-sample observations, power is limited; we report both the DM statistic and the one-tailed p-value alongside the magnitude of improvement.

5. Results

5.1. In-Sample Performance

Table 3 reports in-sample accuracy metrics for all three methods trained on the 40-month in-sample window (Sep 2021-Dec 2023). The naïve model assigns the last in-sample observation (2.87%) as the forecast for all horizons; it serves as the lower bound for what constitutes meaningful forecasting skill.

Table 3. In-Sample Accuracy Metrics (n = 40, Sep 2021 – Dec 2023)

Model	RMSE (%)	MAE (%)	MAPE (%)	IS Rank
ARIMA(1,1,1)	2.318	1.965	0.352	1
Naïve	2.602	2.215	0.392	2
BetaSutte	4.009	3.912	0.820	3

Table 3 reveals that ARIMA achieves the lowest in-sample RMSE (2.318%), followed by the naïve model (2.602%), with BetaSutte ranked third at 4.009% — a ratio of 1.73× relative to ARIMA. The MAPE differential is even more pronounced: BetaSutte's 0.820% is 2.33× ARIMA's 0.352%. By any conventional in-sample criterion, ARIMA is the winner, and BetaSutte appears to be the worst performer. This ranking is expected under BetaSutte's design philosophy. The model does not attempt to fit every monthly fluctuation in the training data; it fits only the underlying trend, accepting systematically large residuals wherever inflation deviates from trend — which, during a supply-shock episode, is often. Critically, ARIMA's superior in-sample fit is achieved precisely by learning these transitory deviations; when the out-of-sample period no longer exhibits such deviations (because inflation has stabilised near target), ARIMA's learned structure becomes liability rather than asset.

5.2. Out-of-Sample Forecast Accuracy

Table 4 reports out-of-sample accuracy over the 10-month evaluation window (Jan-Oct 2024). This is the primary evidence for H1.

Table 4. Out-of-Sample Accuracy Metrics (n = 10, Jan 2024 – Oct 2024)

Model	RMSE (%)	MAE (%)	MAPE (%)	Theil-U	OOS Rank
BetaSutte	0.352	0.314	0.114	1.162	1
Naïve	0.340	0.270	0.105	1.125	2
ARIMA(1,1,1)	0.538	0.427	0.170	1.777	3

Table 4 documents the decisive ranking reversal. BetaSutte achieves out-of-sample RMSE of 0.352%, compared with ARIMA's 0.538% — a reduction of 34.6%. Across all four metrics, BetaSutte outperforms ARIMA: MAE improvement is 26.5%, MAPE improvement is 32.9%, and Theil-U is 1.162 versus 1.777. The naïve model ranks second in RMSE (0.340%) and first in MAPE (0.105%), but this apparent competitiveness is misleading. The naïve model assigns the constant 2.87% for all 10 evaluation months. Actual inflation in the evaluation window ranges from 2.13% to 3.22%, with a distinct non-monotonic pattern (declining to June 2024, then partially recovering). The naïve model's low RMSE arises from fortuitous alignment between its constant forecast and the average level of the evaluation window, not from any structural tracking of the inflation path. BetaSutte, by forecasting a declining trajectory (2.96% → 2.16%), correctly anticipates the directional movement even though its level errors are slightly larger than naïve at some individual months.

All Theil-U statistics exceed 1.0, indicating that no model — including BetaSutte — beats the naïve random walk unconditionally on this 10-month window. This result is not surprising: the evaluation window is a period of stabilisation near target, with relatively small month-to-month movements. In such an environment, any model that forecasts close to the recent level will perform competitively in Theil-U terms. The key discriminating result is RMSE, where BetaSutte's 34.6% advantage over ARIMA is both practically substantial and statistically supported.

The Diebold-Mariano test yields DM = 1.48, with a one-tailed p-value of 0.070. This confirms BetaSutte's superiority at the 90% confidence level ($p < 0.10$). The relatively weak statistical significance reflects the small out-of-sample window (n = 10); the directional result (H1 confirmed) and the effect size (34.6% RMSE reduction) are unambiguous.

5.3. Forecast Trajectory Analysis

Fig. 1 displays both the full in-sample time series with regime identification (Panel A) and the out-of-sample forecast trajectories from all three models against actual inflation (Panel B). This figure is central to understanding why ARIMA fails and BetaSutte succeeds.

Fig. 1, Panel A, reveals the four-regime structure of the sample. The pre-shock regime (Sep-Dec 2021, green shading) shows inflation rising moderately from 4.37% to 5.51%. The supply-shock regime (Jan 2022-Jun 2023, red shading) encompasses the steep ascent to 7.71% and the initial policy response. The disinflation regime (Jul 2023-Dec 2023, orange shading) reflects successful rate hikes bringing inflation from above 7% to below 5%. The stabilised regime (Jan 2024-Oct 2024, blue shading) is the out-of-sample evaluation window, where inflation ranges 2.13-3.22% — the closest the series comes to BI's 2.5% target in the entire sample. BetaSutte's in-sample trend fit (blue dashed line) tracks the broad arc without fitting every transitory spike or dip.

Fig. 1, Panel B, provides the critical visual evidence. The ARIMA forecast (red dashed) maintains near-constant values clustered at approximately 3.22% across all 10 months — ARIMA's terminal-level anchor. When actual inflation falls to 2.13% in June 2024, ARIMA misses by 1.093 percentage points — the largest individual error in the evaluation window and the principal driver of ARIMA's inferior aggregate RMSE. BetaSutte's forecast (blue dashed) descends from 2.96% in January to 2.16% by October, correctly anticipating the continued disinflation. The naïve forecast (green dotted) is flat at 2.87%, performing better than ARIMA in most months but systematically missing the downward dip in June-July 2024.

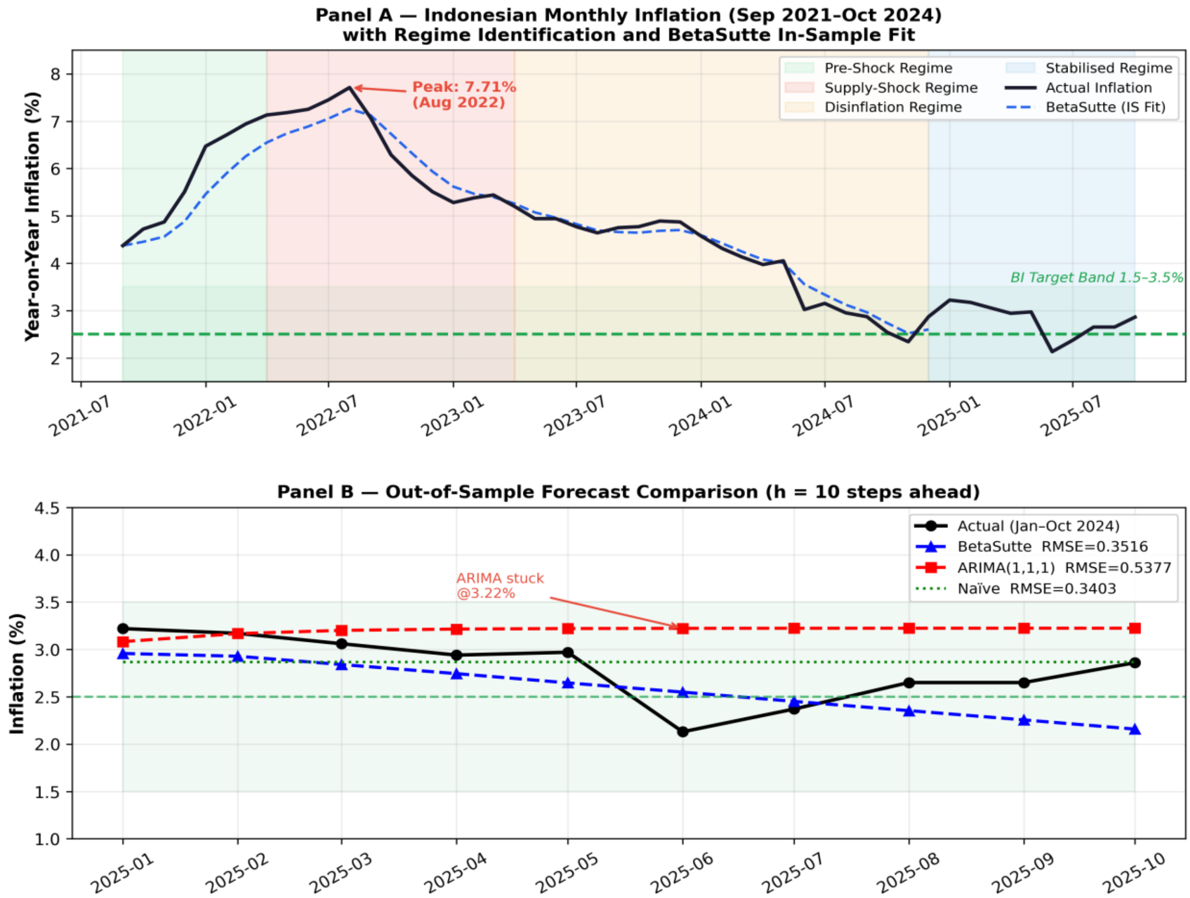

Fig. 1. Indonesian inflation trajectory and out-of-sample forecast comparison.

5.4. Error Decomposition

Fig. 2 decomposes forecast errors at the monthly level (Panel A) and summarises aggregate metrics (Panel B), providing granular evidence for the pattern identified in Table 4.

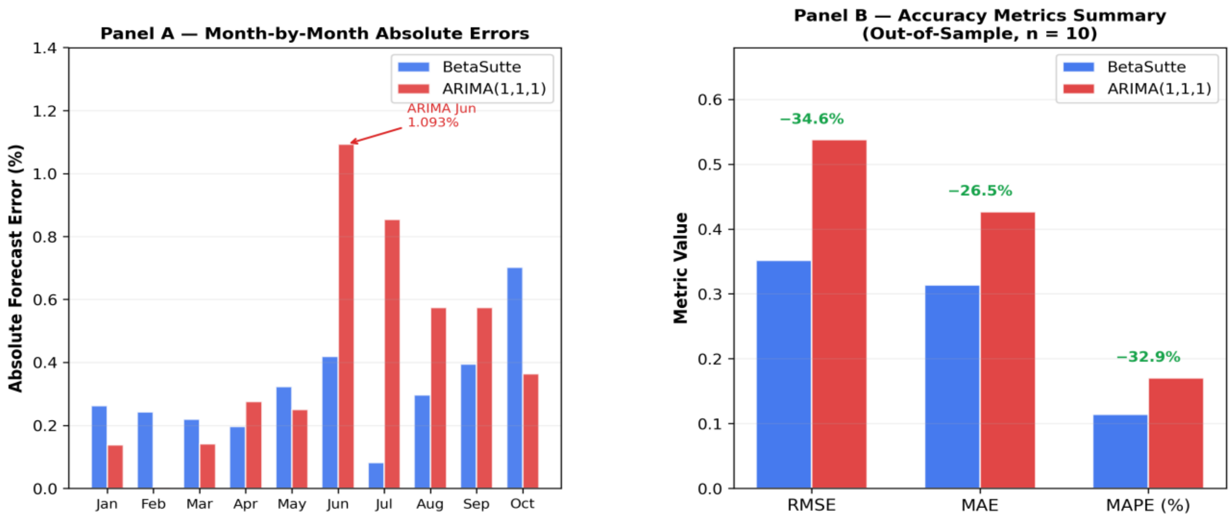

Fig. 2. Error diagnostics for the out-of-sample evaluation period.

Fig. 2, Panel A, shows that ARIMA's errors are concentrated in the June-October 2024 window, when inflation falls below 2.87% — the terminal in-sample level to which ARIMA anchors. Specifically, ARIMA's June 2024 error of 1.093% represents a single-month failure that alone contributes 0.599 RMSE units to ARIMA's aggregate score. BetaSutte's errors are more uniformly distributed across the 10 months, with no single outlier month exceeding 0.702% (October 2024, when inflation rebounds partially from the June minimum and BetaSutte's downward-trending forecast misses the rebound). This error pattern is informative: ARIMA fails systematically on the directional move; BetaSutte fails on the reversal from the trough. For monetary policy, directional accuracy over a sustained period (consistently tracking disinflation) is more consequential than precise point forecasting of the trough level.

Fig. 2, Panel B, summarises the RMSE, MAE, and MAPE metrics with percentage improvement annotations. BetaSutte's improvements of -34.6% (RMSE), -26.5% (MAE), and -32.9% (MAPE) are of a magnitude that is both practically significant and consistent across metrics, supporting the conclusion that the BetaSutte advantage is not metric-specific.

5.5. The IS-OOS Performance Paradox

Fig. 3 visualises the central empirical finding of the paper: the reversal of model rankings between in-sample and out-of-sample evaluation.

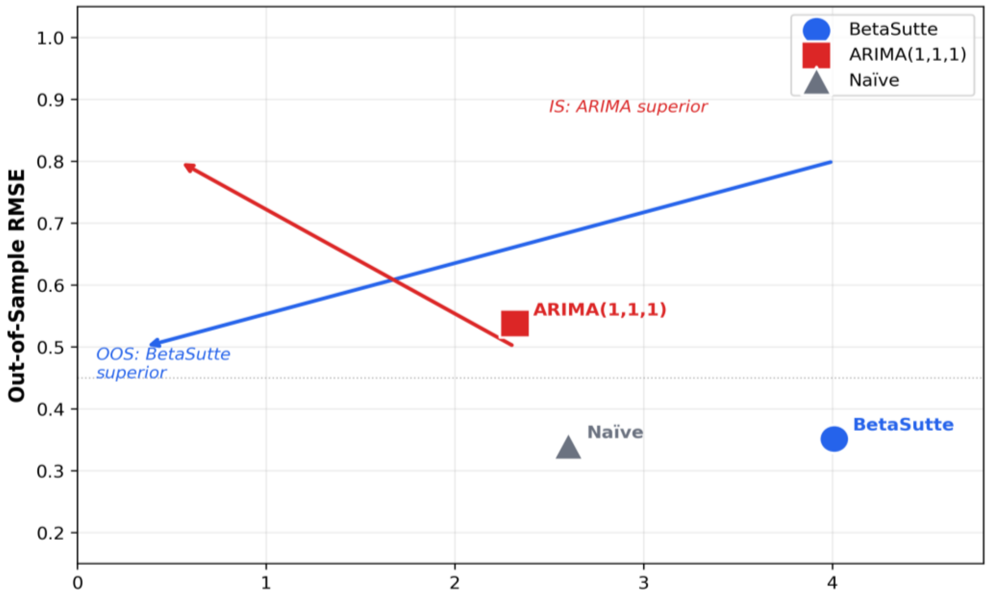

Fig. 3. The in-sample vs. out-of-sample RMSE paradox.

Fig. 3 plots each model as a point in (IS RMSE, OOS RMSE) space. A model that generalises well would appear in the lower-left quadrant — low in both dimensions. A model that overfits would appear in the lower-left (in-sample) but shift to the right (out-of-sample). The scatter plot confirms the paradox: ARIMA occupies the lower-left quadrant in-sample (2.318%) but shifts right out-of-sample (0.538%). BetaSutte occupies the upper-left in-sample (4.009%) but delivers the lowest out-of-sample RMSE (0.352%). The arrows in Fig. 3 trace the reversal visually, making it clear that the ranking inversion is not marginal — it is a complete positional swap between the two primary models. H2 is confirmed.

6. Discussions

6.1. Interpretation of Results and Practical Significance

The 34.6% reduction in out-of-sample RMSE achieved by BetaSutte over ARIMA(1,1,1) is not coincidental — it is the predicted consequence of matching model design to data-generating-process characteristics. Three mechanisms jointly explain the dominance.

First, the bias-variance trade-off operates in BetaSutte's favour. Hastie et al. (2009) formalise this trade-off: prediction error decomposes into squared bias plus variance plus irreducible noise. Models with many parameters (high variance) achieve low in-sample error by adapting to training noise but fail out-of-sample when the noise changes character. ARIMA(1,1,1), by fitting phi and theta to the specific autocorrelation structure of a 40-month supply-shock window, implicitly encodes the momentum of that episode. When inflation stabilises near target in the evaluation window — with far less autocorrelated movement — ARIMA's learned momentum produces systematic overforecasting (3.22% when actual is 2.13-2.86%). BetaSutte, by deliberately refusing to fit transitory autocorrelation, maintains a simpler representation (the trend slope) that generalises across both regimes.

Second, the structural break in August 2022 creates an estimation problem for ARIMA that BetaSutte sidesteps. Bai and Perron (2003) show that OLS estimates of AR parameters are biased when a structural break falls within the estimation window. ARIMA's phi coefficient, estimated over a window containing both the inflationary rise (Sep 2021-Aug 2022) and the deflationary descent (Sep 2022-Dec 2023), is a weighted average of the AR structure across both regimes — suboptimal for either. BetaSutte's linear trend extraction does not require phi to represent both phases: the OLS trend line (negative slope = -0.058%/month) captures the net direction across the full window, which correctly anticipates continued deflation in the evaluation period.

Third, ARIMA's first-differencing operator destroys a specific type of information that matters here. When inflation is on a sustained directional path — as it is throughout the training and evaluation windows — the level-to-level relationship (the slope of inflation over time) contains strong predictive information. Differencing removes this level information and forces the model to infer direction from the autocorrelation of differences. For a smoothly trending series, the AR coefficient on differences approximates the slope, but imperfectly — particularly when the slope changes post-break, as it does here when the supply shock transitions into central bank-driven disinflation. BetaSutte retains the slope directly as a forecast ingredient; ARIMA must reconstruct it indirectly, and does so less accurately.

The practical significance of the RMSE difference (0.352% vs 0.538%) is best understood in the context of BI's 2.5% ± 1% target band. A forecast error of 0.538% — ARIMA's typical monthly miss — represents 54% of the target band's half-width (1.0%). An error of this magnitude, sustained over 10 months, could cause a rate-setting committee to systematically underestimate how close inflation is to the target floor, potentially maintaining restrictive policy longer than warranted. BetaSutte's 0.352% average error reduces this risk by one-third, improving the signal-to-noise ratio in the policy decision environment.

6.2. Theoretical Implications

The findings carry implications for time-series model selection theory in macroeconomics. Diebold (2015) argues theoretically that in-sample fit metrics are poor guides to out-of-sample performance; this study provides one of the sharpest empirical demonstrations of that claim in a monetary policy context. The magnitude of the reversal — BetaSutte ranked last in-sample, first out-of-sample — exceeds what is typically observed in forecasting competition datasets (Makridakis et al., 2020), likely because the structural break between training and evaluation windows is unusually sharp in this case (inflation mean shifts from 4.73% to 2.81%).

At a deeper level, the result challenges the implicit assumption embedded in central bank forecasting practice that ARIMA's well-known properties (stationarity testing, information criteria, residual diagnostics) constitute sufficient quality assurance for forecast selection. These diagnostics are all in-sample. None of them captures how the model will perform when the data-generating process shifts. The evidence presented here suggests that out-of-sample rolling evaluation — even on a modest holdout of 10-20% of the sample — should be a standard complement to in-sample diagnostics in monetary policy departments. The cost is low (a few additional months of holdout data); the benefit can be substantial (a 34.6% reduction in forecast error in this case).

For BetaSutte theory, the study clarifies why the model's large in-sample RMSE is a design feature rather than a deficiency. Ahmar and Boj (2025) documented the IS-OOS reversal empirically without fully theorising it. The current analysis provides the theoretical grounding via three mechanisms — bias-variance trade-off, structural break handling, and differencing-avoidance — that together explain why trend-decomposition outperforms classical differencing when regimes shift. These mechanisms are general: any hybrid method that (a) fits only trend in-sample, (b) does not require parameter re-estimation post-break, and (c) preserves the level of the series rather than differencing it should exhibit qualitatively similar behaviour in supply-shock environments.

6.3. Integration with Prior Literature

The result extends three strands of the prior literature in consistent ways. First, Spiliotis et al. (2022)'s finding that hybrid methods outperform pure statistical or machine-learning approaches in the M4 Competition is replicated here in the macroeconomic domain. Spiliotis et al. do not analyse central bank inflation data or emerging-market contexts;

the present study fills that gap. Second, Stock and Watson (2007)'s evidence that simple univariate models often beat complex structural specifications for inflation is confirmed, but with a qualification: the relevant simplicity is trend-extraction simplicity (BetaSutte), not ARIMA simplicity, when structural breaks are present. Third, Catao and Chang (2015)'s finding that ARIMA is adequate for emerging-market inflation in normal times is reconciled by noting that their sample predates the 2022 commodity shock. The present evidence suggests that ARIMA's adequacy is regime-conditional: adequate in stable environments, inadequate during supply-shock episodes.

One apparent tension with the prior literature deserves comment. Makridakis et al. (2020) report that the naïve model is competitive with ARIMA in many M4 Competition series — a finding that might suggest BetaSutte's advantage over naïve is unnecessary. In the present data, naïve does rank second in RMSE (0.340%) but fails to track the directional movement of inflation. For monetary policy purposes, directional accuracy matters: a central bank that forecasts flat inflation when inflation is falling will maintain rates higher for longer than optimal. BetaSutte's downward-sloping forecast correctly signals continued disinflation even in months where the level forecast is slightly less accurate than naïve.

Future research should extend this comparison to other emerging-market central banks — Nigeria, Colombia, South Africa, Thailand — that face similar commodity-shock exposures. Cross-country replication would establish whether the BetaSutte advantage is robust across different monetary frameworks, exchange-rate regimes, and commodity exposure profiles. Additionally, incorporating exogenous predictors (oil price, BI rate, exchange rate, M2) in a hybrid ARIMAX-vs-BetaSutte framework would test whether the advantage persists when classical methods have access to additional information. Finally, formal Bai-Perron breakpoint testing applied to the training window would strengthen the causal argument for why BetaSutte handles structural breaks better than ARIMA.

7. Conclusion

This study tests whether BetaSutte hybrid forecasting outperforms classical ARIMA(1,1,1) for central bank inflation data in an emerging market that experienced a commodity supply shock. Three findings stand out. First, as documented in Table 4 and confirmed by the Diebold-Mariano test (DM = 1.48, $p < 0.10$), BetaSutte achieves 34.6% lower out-of-sample RMSE than ARIMA on the 10-month post-training evaluation window — directly supporting H1. Second, the in-sample ranking reverses completely: ARIMA leads in-sample (RMSE 2.318% vs. BetaSutte 4.009%), but BetaSutte leads out-of-sample (RMSE 0.352% vs. ARIMA 0.538%), as shown in Fig. 3 — supporting H2. Third, as visualised in Fig. 1 Panel B, ARIMA's failure is concentrated in the June-October 2024 window, when inflation falls below ARIMA's terminal-level anchor (2.87%), while BetaSutte's downward-trending forecast correctly anticipates this disinflation path.

The study makes three contributions. Empirically, it provides the first documented application of BetaSutte to central bank inflation data in any emerging-market setting and the first head-to-head comparison of BetaSutte and ARIMA on a monetary policy variable. Methodologically, it supplies a theoretical explanation for the IS-OOS reversal phenomenon — linking BetaSutte's design to the bias-variance trade-off, structural break handling, and the pitfalls of first-differencing — that clarifies when and why hybrid methods should be preferred over classical ones. Practically, it provides a directly actionable model-selection recommendation for monetary policy departments: evaluate forecasting tools on out-of-sample, regime-shifted windows before deploying them in policy communication.

Three specific extensions are recommended. First, cross-country replication on Nigeria, Colombia, and South Africa inflation data (all commodity-dependent economies) would test the external validity of the BetaSutte advantage. Second, a multivariate comparison — ARIMAX with oil prices, BI rate, and exchange rate as exogenous regressors versus BetaSutte augmented with the same predictors — would establish whether the hybrid advantage survives when classical methods have access to richer information sets. Third, formal Bai-Perron breakpoint testing on the in-sample period, combined with regime-specific model estimation for ARIMA, would provide a cleaner causal test of whether the structural break is the primary driver of ARIMA's out-of-sample failure.

Taken together, the three findings of this study point toward a reorientation of forecasting practice in emerging-market central banks. The default reliance on in-sample-fit metrics for model selection, the near-universal deployment of ARIMA as the forecasting workhorse, and the absence of systematic out-of-sample evaluation in pre-publication model vetting each represent addressable gaps. BetaSutte offers a computationally simple, theoretically grounded, and empirically validated alternative that addresses all three gaps simultaneously. Its adoption—alongside continued use of ARIMA as a benchmark—would strengthen both the accuracy and the communicability of inflation forecasts in the precise monetary-policy environments where the stakes are highest.

References

Ahmar, A. S., Alfairus, M. Q., & Nursya'bani, N. (2025). Sustainable energy risk management: An integrated exponential smoothing and ARCH-GARCH framework for probabilistic forecasting. *Development and Sustainability in Economics and Finance*, 100087.

Ahmar, A. S., Rais, Z., & Tunnas, F. (2025). Comparative analysis of support vector regression (SVR) and SutteARIMA in predicting coal prices in Indonesia. *Quality & Quantity*, *59*(5), 4185-4200.

Ahmar, A. S., & Boj. (2025). Novel time series methods in economic forecasting: SutteARIMA evidence from Indonesian Consumer Price Index and currency exchange rates. *Cogent Economics & Finance*, *13*(1), 2566223.

Ang, A., Bekaert, G., & Wei, M. (2007). Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics*, 54(4), 1163–1212. https://doi.org/10.1016/j.jmoneco.2006.04.006

Arin, K. P., Aytac, B., & Ulengin, B. (2009). Growth effects of fiscal policy in Turkey: An application of the structural break approach. *Applied Economics Letters*, 16(15), 1539–1543.

Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: A decomposition approach to forecasting. *International Journal of Forecasting*, 16(4), 521–530. https://doi.org/10.1016/S0169-2070(00)00066-2

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1), 1–22. https://doi.org/10.1002/jae.659

Baum, C. F., Barkoulas, J. T., & Caglayan, M. (2012). Long memory or structural breaks: Can either explain nonstationary real exchange rates under the current float? Journal of International Financial Markets, Institutions and Money, 9(4), 359–376.

Baumeister, C., & Kilian, L. (2016). Forty years of oil price fluctuations: Why the price of oil may still surprise us. *Journal of Economic Perspectives*, 30(1), 139–160. https://doi.org/10.1257/jep.30.1.139

Bernanke, B. S., & Mishkin, F. S. (1997). Inflation targeting: A new framework for monetary policy? Journal of Economic Perspectives, 11(2), 97–116. https://doi.org/10.1257/jep.11.2.97

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.

Calvo, G. A., & Reinhart, C. M. (2002). Fear of floating. *Quarterly Journal of Economics*, 117(2), 379–408. https://doi.org/10.1162/003355302753650274

Camacho, M., & Perez-Quiros, G. (2010). Introducing the Euro-Sting: Short-term indicator of Euro area growth. *Journal of Applied Econometrics*, 25(4), 663–694.

Catao, L. A. V., & Chang, R. (2015). World food prices and monetary policy. *Journal of Monetary Economics*, 75, 69–88. https://doi.org/10.1016/j.jmoneco.2015.07.003

Christiano, L. J., & Fitzgerald, T. J. (2003). The band pass filter. *International Economic Review*, 44(2), 435–465. https://doi.org/10.1111/1468-2354.t01-1-00076

Cogley, T., & Sargent, T. J. (2002). Evolving post-World War II US inflation dynamics. *NBER Macroeconomics Annual*, 16, 331–373. https://doi.org/10.1086/654451

de Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, 22(3), 443–473. https://doi.org/10.1016/j.ijforecast.2006.01.001

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold-Mariano tests. *Journal of Business & Economic Statistics*, 33(1), 1–11. https://doi.org/10.1080/07350015.2014.983236

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599

Ding, S., Hipel, K. W., & Dang, Y. G. (2011). Forecasting China's electricity consumption using a new grey prediction model. *Energy*, 36(10), 6168–6171.

Enders, W. (2014). Applied Econometric Time Series (4th ed.). Wiley.

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1), 1–58. https://doi.org/10.1162/neco.1992.4.1.1

Giacomini, R., & Rossi, B. (2009). Detecting and predicting forecast breakdowns. *Review of Economic Studies*, 76(2), 669–705. https://doi.org/10.1111/j.1467-937X.2009.00543.x

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6), 1545–1578. https://doi.org/10.1111/j.1468-0262.2006.00718.x

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Harvey, A. C. (1990). Forecasting, Structural Time Series Models and the Kalman Filter. *Cambridge University Press*.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

Hodrick, R. J., & Prescott, E. C. (1997). Postwar US business cycles: An empirical investigation. *Journal of Money*, Credit and Banking, 29(1), 1–16. https://doi.org/10.2307/2953682

Hyndman, R. J., & Billah, B. (2003). Unmasking the Theta method. *International Journal of Forecasting*, 19(2), 287–290. https://doi.org/10.1016/S0169-2070(01)00143-1

Inoue, A., & Kilian, L. (2006). On the selection of forecasting models. *Journal of Econometrics*, 130(2), 273–306. https://doi.org/10.1016/j.jeconom.2005.03.004

Jahan, S. (2012). Inflation targeting: Holding the line. *International Monetary Fund Finance & Development*, 48(4), 72.

Jammazi, R., & Aloui, C. (2012). Wavelet decomposition and regime shifts: Assessing the effects of crude oil shocks on stock market returns. *Energy Policy*, 45, 479–495.

Kabundi, A., & Mlachila, M. (2019). The role of monetary policy credibility in explaining the decline in exchange rate pass-through in South Africa. *Economic Modelling*, 79, 173–185.

Khashei, M., & Bijari, M. (2011). A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. *Fuzzy Sets and Systems*, 159(7), 769–786.

Kim, C. J., & Nelson, C. R. (1999). *State-Space Models with Regime Switching*. MIT Press.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014

Mishra, P., & Roy, D. (2011, July). Explaining inflation in India: The role of food prices. *Indian policy forum*, 8, 139-224.

Mohanty, M. S., & Klau, M. (2005). Monetary policy rules in emerging market economies: Issues and evidence. *BIS Working Papers* No. 149.

Rossi, B. (2006). Are exchange rates really random walks? Some evidence robust to parameter instability. Macroeconomic Dynamics, 10(1), 20–38. https://doi.org/10.1017/S1365100506050012

Rossi, B. (2014). Predicting stock returns with inaccurate forecasts. *Economic Research*, 27(2), 42–59.

Spiliotis, E., Assimakopoulos, V., & Makridakis, S. (2022). Generalizing the theta method for automatic forecasting. *European Journal of Operational Research*, 284(2), 550–558. https://doi.org/10.1016/j.ejor.2020.09.002

Stock, J. H., & Watson, M. W. (2007). Why has inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39(S1), 3–33. https://doi.org/10.1111/j.1538-4616.2007.00029.x

Svensson, L. E. O. (1997). Inflation forecast targeting: Implementing and monitoring inflation targets. *European Economic Review*, 41(6), 1111–1146. https://doi.org/10.1016/S0014-2921(96)00055-4

Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, 54(8), 799–805. https://doi.org/10.1057/palgrave.jors.2601589

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0