International Journal of Management Science and Information Technology
IJMSIT

E-ISSN: 2774-5694 P-ISSN: 2776-7388 OPEN ACCESS

Volume 6 (2), July-December 2026, 1832-1844
DOI: https://doi.org/10.35870/ijmsit.v6i2.8180

Structural Change Detection and Forecasting of Indonesia's Tourism-Related Output Using PELT and SARIMA Intervention Models

Haves Qausar [1*], Zata Hasyyati [2]
[1*] Mathematics Education Study Program, Faculty of Teacher Training and Education, Universitas Malikussaleh, North Aceh Regency, Aceh Province, Indonesia
[2] Development Economics Study Program, Faculty of Economics and Business, Universitas Malikussaleh, North Aceh Regency, Aceh Province, Indonesia

Email: haves@unimal.ac.id [1*]; zatahasyyati@unimal.ac.id [2]

Article history:
Received July 24, 2026
Revised August 7, 2026
Accepted August 13, 2026

Abstract
Quarterly tourism-related output is seasonal and highly vulnerable to structural disruptions, so models fitted to stable historical patterns may fail after shocks. This study detects changepoints and compares forecasting models for Indonesia's real GDP in Accommodation and Food Service Activities, used as a proxy for tourism-related economic output. The dataset comprised 50 quarterly observations at constant 2010 prices from 2013Q1 to 2025Q2. PELT was applied to detrended, seasonally adjusted log-training residuals. Ordinary SARIMA, SARIMA models with pulse and step interventions, six exponential-smoothing specifications, and a seasonal-naive benchmark were screened using training AICc, 16 rolling one-step validation origins, and an untouched ten-quarter fixed test. PELT identified 2017Q2, 2020Q2, and 2022Q1 as changepoints, with 2020Q2 remaining the most robust under stronger penalties. SES achieved the smallest rolling-validation RMSE (6,283.9), while the pulse-plus-temporary-step SARIMA intervention was best within its family (RMSE 6,917.4). On the fixed 2023Q1-2025Q2 test, SARIMA $(0,1,1) \times (0,0,0)_4$ performed best overall, with MAE of Rp5,301.9 billion, RMSE of Rp6,131.6 billion, sMAPE of 5.23%, and MASE of 1.016. For the selected baseline SARIMA, residual autocorrelation and ARCH effects were not significant; the intervention model, however, retained significant residual autocorrelation. Residuals were non-normal across the retained models, and the baseline SARIMA underpredicted the later recovery. PELT provided structural diagnosis, but explicit intervention effects did not guarantee superior holdout forecasts. Parsimonious SARIMA provides the most defensible baseline, supplemented by changepoint monitoring and scenario-based intervention analysis.

Keywords:
Changepoint detection; PELT; SARIMA intervention; Tourism forecasting; Indonesia.

1. INTRODUCTION

Tourism-related activities connect visitor expenditure with accommodation, food services, transport, employment, and local supply chains, making timely forecasts relevant to both public policy and business planning. The sector is also highly exposed to mobility restrictions, income shocks, and changes in traveller confidence. This combination of economic importance and shock sensitivity became especially visible during and after the COVID-19 pandemic. At the global level, international tourism returned to approximately 99% of its 2019 level in 2024, with an estimated 1.4 billion international arrivals (UN Tourism, 2025). Aggregate recovery, however, does not imply that the underlying quarterly path is stable. Tourism agencies and hospitality businesses still require forecasts that distinguish recurring seasonal movement from temporary disruptions and persistent changes in the data-generating process.

In Indonesia, the need for structurally aware forecasting is evident in the trajectory of the national economy and tourism-related industries. Statistics Indonesia reported that real GDP contracted by 5.32% year-on-year and 4.19% quarter-to-quarter in 2020Q2 (Badan Pusat Statistik [BPS], 2020a), whereas annual economic growth recovered to 5.31% in 2022 from 3.70% in 2021 (BPS, 2023). For quarterly monitoring, this study uses real GDP at constant 2010 prices for Category I, Accommodation and Food Service Activities, as an operational indicator of tourism-related economic output. The series is consistently reported in BPS quarterly GDP publications (BPS, 2024, 2025b). Because this industrial category also includes food-service consumption by residents and other non-tourism activity, it is treated as a proxy rather than as Indonesia's complete Tourism Direct Gross Domestic Product. Even with this limitation, it offers a policy-relevant measure of real production in two industries that are closely exposed to visitor and mobility dynamics.

Forecasting such a series is statistically demanding. Quarterly output can contain trend, seasonality, serial dependence, and abrupt level changes, while the available sample is short relative to the number of components that could be estimated. Seasonal autoregressive integrated moving average (SARIMA) models provide a parsimonious representation of stochastic trend and seasonal dependence (Box et al., 2015), whereas exponential-smoothing models provide a complementary state-space description of level, trend, and seasonality (Hyndman et al., 2002). Both families nevertheless rely on historical regularities that can be distorted by an exceptional interruption. Intervention analysis addresses this issue by embedding pulse or step effects in a dynamic regression whose errors retain an ARIMA structure, thereby separating an event-related effect from dependent noise (Box & Tiao, 1975).

Previous tourism-forecasting studies support the usefulness of these models but also illustrate the unresolved difficulties created by the pandemic. Zhang et al. (2021) combined econometric and judgemental methods to construct alternative recovery paths for Hong Kong, demonstrating that forecasts based only on pre-pandemic regularities could become obsolete after a major shock. In the Indonesian context, Rianda and Usman (2023) modelled international visitor arrivals using ARIMAX and intervention specifications and found the intervention model more accurate while also quantifying the pandemic effect. Fahmiyah et al. (2024) compared SARIMA and Holt-Winters models for monthly foreign-tourist visits to Indonesia and reported lower RMSE and AIC values for SARIMA. These studies establish that seasonal models, external information, and intervention variables can improve tourism-demand analysis, but their response variable is visitor arrivals rather than quarterly real output from accommodation and food services.

A second strand of literature concerns the timing of structural change. The Pruned Exact Linear Time (PELT) algorithm detects multiple changepoints by minimizing a penalized segmentation cost and, under mild conditions, achieves computational cost that is linear in the number of observations (Killick et al., 2012). This makes PELT useful for identifying candidate regime boundaries without fixing every date in advance. Changepoint detection and intervention modelling nevertheless answer different questions: the former locates distributional changes, whereas the latter estimates specified shock patterns while accounting for serially correlated errors. Recent work on interrupted time-series forecasting also emphasizes that disruptions create a distinct forecasting problem and that post-interruption performance must be assessed rather than inferred from pre-interruption fit alone (Hyndman & Rostami-Tabar, 2025).

Taken together, the cited studies leave three specific gaps for the present application. First, Indonesian tourism forecasting has largely emphasized monthly visitor counts, leaving quarterly real production in accommodation and food services less examined as a forecasting target. Second, intervention dates are commonly defined directly from known events; this is substantively reasonable but does not show whether the observed series contains additional or differently timed regime changes. Third, the literature reviewed here does not provide an integrated comparison in which training-only changepoint detection informs, but does not dictate, competing pulse and step interventions, after which SARIMA, SARIMA-intervention, exponential-smoothing, and seasonal-naive models are evaluated through rolling-origin validation and a completely untouched fixed test. The last gap is important because in-sample fit or a single model-selection criterion cannot substitute for genuine out-of-sample evidence (Tashman, 2000).

The novelty of this study therefore lies in methodological integration and validation design, not in claiming that PELT, SARIMA, or intervention analysis is itself new. The proposed PELT-SARIMA-intervention framework first removes deterministic trend and quarterly effects from the log-transformed training series, applies PELT with penalty-sensitivity analysis, and treats persistent changepoints as empirical evidence about possible regimes. It then estimates alternative pulse, temporary-step, permanent-step, and combined intervention structures within SARIMA-error models. Ordinary SARIMA, six exponential-smoothing/Holt-Winters specifications, and a seasonal-naive benchmark are retained as comparators. Model choices are made using only the 2013Q1-2022Q4 training period and rolling validation, while the ten quarters from 2023Q1 to 2025Q2 remain untouched until final evaluation. This separation prevents detected breaks or forecasts errors in the test period from influencing model design.

Accordingly, this study aims to: identify robust changepoints in Indonesia's quarterly real GDP for Accommodation and Food Service Activities; determine whether explicitly modelled interventions improve forecasting relative to an ordinary SARIMA process; compare the retained SARIMA and SARIMA-intervention specifications with exponential smoothing and a seasonal-naive benchmark; and evaluate predictive performance on a fixed post-training test period using MAE, RMSE, sMAPE, and MASE. The

analysis also seeks to interpret the resulting regimes, coefficients, diagnostics, and forecast errors from mathematical, economic, and managerial perspectives.

The study contributes in three ways. Mathematically, it demonstrates a reproducible sequence that links distributional changepoint evidence to dynamic intervention models without treating an algorithmically detected date as proof of causality. Empirically, it shifts attention from tourist-arrival counts to a quarterly constant-price production indicator spanning the pre-pandemic, disruption, and recovery periods. Managerially, it provides a transparent forecasting benchmark for tourism agencies, hospitality operators, and regional planners that can be updated when new observations become available. The remainder of the paper is organized as follows. Section 2 explains the data, transformations, PELT procedure, forecasting models, validation design, diagnostics, and accuracy measures. Section 3 presents and discusses the descriptive results, changepoints, selected models, and fixed-test forecasts. Section 4 concludes with the main implications, limitations, and recommendations.

2. RESEARCH METHOD

2.1. Research Design, Variable, and Data Source

This study employed a quantitative, retrospective time-series design to model structural change and forecast quarterly tourism-related economic output in Indonesia. The response variable, Y_t, was the quarterly gross domestic product (GDP) at constant 2010 prices for Category I, Accommodation and Food Service Activities, expressed in billions of Indonesian rupiah. This category was used as an operational proxy for tourism-related output; it should not be interpreted as Indonesia's complete Tourism Direct Gross Domestic Product. The use of a real-price series removed the direct effect of general price changes and made intertemporal comparisons meaningful.

The dataset consisted of 50 consecutive quarterly observations from 2013Q1 to 2025Q2. The values were compiled from five overlapping quarterly GDP publications issued by Statistics Indonesia (BPS), 2017, 2019, 2020b, 2024, 2025b). Only observations with the same industrial classification, constant-price base year, unit, and quarterly frequency were combined. Overlapping years were reconciled before analysis, and no interpolation or imputation was required.

Table 1. Data partition and analytical purpose

Partition	Period	n	Purpose
Full series	2013Q1–2025Q2	50	Descriptive presentation and final reporting
Training	2013Q1–2022Q4	40	Stationarity tests, PELT, estimation, and model selection
Rolling validation	2019Q1–2022Q4	16 origins	One-step-ahead comparison after a 24-quarter initial window
Fixed test	2023Q1–2025Q2	10	Final multi-step out-of-sample evaluation

The split was chronological to preserve temporal ordering and prevent look-ahead bias. The first 40 observations formed the training set. The final 10 observations were held out and were not used for changepoint detection, order selection, intervention selection, parameter tuning, or rolling validation. This design follows the principle that forecasting models must be evaluated using information that would have been available at each forecast origin (Tashman, 2000).

2.2. Data Transformation and Stationarity Assessment

The level series was transformed using the natural logarithm, $z_t = \ln(Y_t)$, to stabilize relative variation and allow intervention coefficients to be interpreted approximately as proportional changes. Descriptive analysis was conducted on Y_t, z_t, quarter-to-quarter log growth, and year-on-year log growth. The two growth measures were defined as follows:

$$g_t^{\mathrm{QoQ}} = 100[\ln(Y_t) - \ln(Y_{t-1})] \quad (1)$$

$$g_t^{\mathrm{YoY}} = 100[\ln(Y_t) - \ln(Y_{t-4})] \quad (2)$$

Stationarity was assessed on the training sample using the augmented Dickey–Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests, which have complementary null hypotheses (Dickey & Fuller, 1979; Kwiatkowski et al., 1992). The ADF regression with an intercept was:

$$\Delta x_t = \alpha + \gamma x_{t-1} + \sum_{i=1}^{p} \delta_i \Delta x_{t-i} + u_t \quad (3)$$

Where the ADF null hypothesis is H_0: $\gamma = 0$ (unit root). The lag p was selected by the Akaike information criterion over a maximum of four lags. The KPSS statistic for level stationarity was:

$$\text{KPSS} = \frac{T^{-2}\sum_{t=1}^{T} S_t^2}{\hat{\sigma}_{LR}^2} \quad (4)$$

Where S_t is the partial sum of residuals and $\hat{\sigma}^2$LR is the Bartlett-kernel long-run variance estimate. The bandwidth was $\lfloor 12(T/100)^{1/4} \rfloor$. Decisions were made at the 5% level using an ADF critical value of −2.93 for the intercept specification and a KPSS critical value of 0.463. Joint evidence from ADF and KPSS was used to determine the differencing order rather than relying on one test alone.

2.3. PELT Changepoint Detection

Changepoints were detected exclusively in the training set. Because a trending log-level series can produce artificial mean shifts, z_t was first regressed on an intercept, a linear time trend, and three quarterly indicators. The standardized residual r_t from the following deterministic regression was used as the PELT input:

$$z_t = a + bt + \gamma_2 D_{2,t} + \gamma_3 D_{3,t} + \gamma_4 D_{4,t} + r_t \quad (5)$$

The Pruned Exact Linear Time (PELT) criterion identifies a partition that minimizes the sum of within-segment costs and a penalty for each additional changepoint (Killick et al., 2012):

$$\min_{m,\boldsymbol{\tau}} \left\{ \sum_{j=0}^{m} C\left(\tau_j + 1, \tau_{j+1}\right) + \beta m \right\} \quad (6)$$

With $\tau_0 = 0$, $\tau_{m+1} = n$, and segment cost $C(a{:}b) = \Sigma[t=a,\ldots,b](r_t - \bar{r}[a{:}b])^2$. A minimum segment length of four quarters was imposed. Because n = 40 was small, the penalized PELT objective was also solved by exact dynamic programming, yielding the global minimum for each penalty. The primary penalty was $\beta = 1.5 \ln(n)$. Sensitivity was evaluated at c ln(n), where $c \in \{0.5, 0.75, 1.0, 1.25, 1.5, 2.0, 2.5, 3.0, 4.0\}$. A break was considered robust when it persisted over several adjacent penalty values. This sensitivity rule reduced dependence on a single arbitrary penalty.

The changepoint dates informed, but did not automatically determine, the intervention specification. Four deterministic intervention structures were examined: a pulse at 2020Q2, a temporary step from 2020Q2 through 2021Q4, a permanent step beginning in 2020Q2, and a combined pulse plus temporary step. Thus, both the abrupt contraction and the multi-quarter disruption associated with COVID-19 could be represented.

2.4. Forecasting Models

The principal model family was seasonal autoregressive integrated moving average (SARIMA). For the log-transformed response z_t and quarterly seasonal period s = 4, the general model was:

$$\Phi(B^4)\phi(B)(1-B)^d(1-B^4)^D z_t = c + \Theta(B^4)\theta(B)\varepsilon_t \quad (7)$$

Where B is the backshift operator; φ(B) and θ(B) are the non-seasonal autoregressive and moving-average polynomials of orders p and q; Φ(B⁴) and Θ(B⁴) are the seasonal polynomials of orders P and Q; d and D are the non-seasonal and seasonal differencing orders; and ε_t is a zero-mean innovation (Box et al., 2015). The stationarity assessment fixed d = 1. Candidate orders used $p, q \in \{0, 1, 2\}$, $P, Q \in \{0, 1\}$, $D \in \{0, 1\}$, and s = 4, subject to $1 \leq p + q + P + Q \leq 4$. This produced 60 SARIMA candidates while limiting over-parameterization in a 40-observation training sample. The intervention model was formulated as dynamic regression with SARIMA errors (Box & Tiao, 1975):

$$z_t = \mu + \beta^P P_t + \beta^T T_t + \beta^S S_t + n_t, \quad n_t \sim \text{SARIMA}(p,d,q)(P,D,Q)_4 \quad (8)$$

Where $P_t = 1$ only at 2020Q2; $T_t = 1$ from 2020Q2 to 2021Q4; and $S_t = 1$ from 2020Q2 onward. The single pulse, temporary step, permanent step, and pulse-plus-temporary structures were fitted to the three SARIMA orders with the smallest AICc, producing 12 intervention candidates. The six intervention candidates with the smallest AICc were advanced to rolling validation.

SARIMA and intervention parameters were estimated by Gaussian conditional sum of squares (CSS). Numerical optimization minimized $n_e \ln(\text{SSE}/n_e)$, where n_e is the effective residual count after differencing and initialization. AR and MA coefficients used bounded parameterization, convergence was required, and three starting values were evaluated to reduce sensitivity to local optima. Log-scale point forecasts were returned to the original scale using $\exp(\hat{z}_t)$; therefore, they are median-scale forecasts and no lognormal mean-bias correction was applied.

Exponential smoothing models were included as non-ARIMA comparators because they represent level, trend, damping, and seasonal evolution directly (Hyndman et al., 2002). The candidates were simple exponential smoothing (SES), Holt's linear trend, damped Holt, additive Holt–Winters, damped additive Holt–Winters, and multiplicative Holt–Winters. The SES update was

$$\ell_t = \alpha Y_t + (1-\alpha)\ell_{t-1}, \quad \hat{Y}_{t+h|t} = \ell_t \tag{9}$$

For additive Holt–Winters, the level, trend, seasonal, and forecast recursions were

$$\ell_t = \alpha(Y_t - s_{t-4}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \tag{10}$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)\phi b_{t-1} \tag{11}$$

$$s_t = \gamma(Y_t - \ell_t) + (1-\gamma)s_{t-4} \tag{12}$$

$$\hat{Y}_{t+h|t} = \ell_t + \left(\sum_{i=1}^{h} \phi^i\right) b_t + s_{t+h-4(k+1)} \tag{13}$$

Where $0 < \alpha, \beta, \gamma < 1$, $0.80 < \varphi \leq 1$, and k is the integer that selects the appropriate seasonal index.

Table 2. Candidate model families

Family	Candidate specification	Estimated candidates	Validation rule
SARIMA	p,q=0–2; P,Q=0–1; d=1; D=0/1; s=4; complexity ≤4	60	Six smallest AICc
SARIMA intervention	Three leading SARIMA orders × four intervention structures	12	Six smallest AICc
ETS/Holt–Winters	SES, Holt, damped Holt, additive, damped additive, multiplicative	6	All candidates
Seasonal-naïve	Quarterly lag-four repetitions	1	Benchmark

The multiplicative seasonal form replaced seasonal subtraction and addition with division and multiplication. Initial level was the mean of the first four quarters, initial trend was the difference between the means of the first two seasonal cycles divided by four, and initial seasonal indices were derived from the first cycle. Smoothing parameters were estimated by minimizing the sum of squared one-step errors.
A seasonal-naïve benchmark was also evaluated. Its forecast repeats the last observed value from the same quarter:

$$\hat{Y}_{t+h|t} = Y_{t+h-4(k+1)}. \tag{14}$$

2.5. Model Selection, Rolling Validation, and Test Evaluation

AICc was used to screen models only within the same model family and response scale. It was not used to rank a log-SARIMA model against an original-scale ETS model. For k estimated parameters and n_e effective residuals,

$$\text{AICc} = \text{AIC} + \frac{2k(k+1)}{n_e - k - 1} \tag{15}$$

The correction is particularly important with only 40 training observations and discourages models whose apparent fit is obtained through excessive parameterization (Hurvich & Tsai, 1989). Together with the restricted candidate grid, it operationalized the study's deliberate preference for parsimonious specifications under limited sample information. After AICc screening, one-step-ahead rolling-origin validation was conducted from 2019Q1 to 2022Q4. The initial estimation window contained 24 quarters (2013Q1–2018Q4). At each of the 16 origins, the model was re-estimated using all observations available at that origin, and the following quarter was forecast. Within each family, the candidate with the smallest rolling RMSE was retained. The selected representative from each family was then re-estimated once using the complete 40-quarter training set and used to generate a fixed ten-quarter forecast for 2023Q1–2025Q2. This final test was strictly confirmatory: test observations were revealed only after all model and intervention choices had been frozen. The complete leakage-free analytical procedure was as follows:

a. Compile and reconcile the quarterly BPS series while preserving the original observations.
b. Fix the training, rolling-validation, and test periods before statistical analysis.
c. Calculate the log and growth transformations and assess stationarity using training data only.
d. Remove deterministic trend and quarterly seasonality, then run the PELT penalty-sensitivity analysis on training residuals.
e. Estimate 60 SARIMA, 12 SARIMA-intervention, six ETS/Holt–Winters, and one seasonal-naïve candidate.
f. Screen candidates by AICc within family and compare the retained candidates by rolling-origin RMSE.
g. Refit each family winner on the full training sample and freeze all specifications.
h. Generate fixed multi-step test forecasts and calculate MAE, RMSE, sMAPE, and MASE.
i. Conduct residual diagnostics and report predictive accuracy together with model adequacy.

2.6. Residual Diagnostics

Residual adequacy was assessed for the selected SARIMA, SARIMA-intervention, and ETS representatives. The Ljung–Box portmanteau test evaluated residual autocorrelation at lags 4 and 8 (Ljung & Box, 1978):

$$Q(h) = n(n+2)\sum_{j=1}^{h}\frac{\hat{\rho}_j^2}{n-j} \tag{16}$$

Where $\hat{\rho}_j$ is the residual autocorrelation at lag j. A p-value of at least 0.05 indicated that the remaining serial dependence was not statistically significant. Residual normality was examined using the Jarque–Bera statistic (Jarque & Bera, 1980):

$$\mathrm{JB} = \frac{n}{6}\left[S^2 + \frac{(K-3)^2}{4}\right] \tag{17}$$

Where S and K are residual skewness and kurtosis. Conditional heteroskedasticity was assessed with Engle's ARCH-LM test using up to four lags of squared residuals (Engle, 1982). The auxiliary regression LM statistic, nR^2, was compared with a chi-square distribution. These diagnostics were interpreted jointly: white-noise residuals were required for model adequacy, whereas non-normality was treated as a warning for Gaussian prediction intervals rather than as automatic evidence of poor point forecasts. The primary evaluation therefore focused on point-forecast accuracy. Conventional Gaussian prediction intervals were not reported because the extreme, heavy-tailed pandemic residuals could make their nominal coverage unreliable; residual-bootstrap or simulation-based intervals are recommended for subsequent extensions.

2.7. Forecast Accuracy Measures

For an evaluation sample of H forecasts, let $e_t = Y_t - \hat{Y}_t$. Accuracy was summarized using mean absolute error (MAE), root mean squared error (RMSE), symmetric mean absolute percentage error (sMAPE), and mean absolute scaled error (MASE). Using multiple measures reduces dependence on the loss function implicit in a single statistic (Hyndman & Koehler, 2006).

$$\mathrm{MAE} = \frac{1}{H}\sum_{t=1}^{H}|e_t| \tag{18}$$

$$\mathrm{RMSE} = \left[\frac{1}{H}\sum_{t=1}^{H}e_t^2\right]^{1/2} \tag{19}$$

$$\mathrm{sMAPE} = \frac{100}{H}\sum_{t=1}^{H}\frac{2|e_t|}{|Y_t|+|\hat{Y}_t|} \tag{20}$$

$$\mathrm{MASE} = \frac{\frac{1}{H}\sum_{t=1}^{H}|e_t|}{\frac{1}{n-4}\sum_{t=5}^{n}|Y_t - Y_{t-4}|} \tag{21}$$

MAE and RMSE retain the original unit of billions of rupiah; RMSE penalizes large errors more strongly. sMAPE is scale-free and bounded under the stated definition. MASE scales the test MAE by the mean absolute seasonal-naïve error calculated only from the training sample. A MASE below one indicates improvement over the average in-sample seasonal-naïve scale.

2.8. Computational Reproducibility and Research Ethics

All computations were executed in a reproducible Python 3.12 workflow using double-precision numerical operations (NumPy 2.3.5 and SciPy 1.17.0); figures were prepared with Matplotlib 3.10.8. Every rolling origin refitted the candidate model rather than reusing parameters from the full sample. Deterministic intervention indicators were generated directly from quarterly date labels. The analysis workbook retains the original observations, transformations, candidate models, forecasts, diagnostics, and source URLs so that each reported value can be audited.

The study used publicly available, aggregated macroeconomic statistics and involved no human participants, personal identifiers, surveys, or experimental treatment. Consequently, informed consent and participant-level ethical approval were not applicable. The principal methodological limitations are the short quarterly sample, the exceptional magnitude of the pandemic shock, the use of Category I GDP as a proxy rather than complete tourism direct GDP, and the absence of external predictors in the univariate forecasting comparison. These limitations were addressed through AICc, parsimonious candidate grids, rolling validation, a fixed holdout test, explicit intervention models, and sensitivity analysis for changepoint penalties.

3. RESULTS AND DISCUSSION

3.1. Descriptive Patterns and Stationarity

The reconciled BPS series (BPS, 2024, 2025b) shows that the quarterly real GDP proxy increased from Rp59,543.3 billion in 2013Q1 to Rp112,298.3 billion in 2025Q2, an 88.6% cumulative increase. This long-run rise was interrupted by an exceptional contraction in 2020Q2 and an uneven recovery through 2021. The mean of the fixed-test period was Rp101,587.9 billion, 37.4% above the training mean, while its coefficient of variation was lower (6.29% versus 11.42%). The lower relative variation does not imply a stable level; rather, the test period occupied a higher and continuing recovery regime.

Table 3. Descriptive statistics of quarterly real GDP

Sample	n	Mean	Median	SD	Minimum	Maximum
Full sample	50	79,469.2	77,466.0	13,748.6	59,543.3	112,298.3
Training	40	73,939.5	73,923.7	8,442.0	59,543.3	93,352.2
Fixed test	10	101,587.9	102,320.8	6,388.2	91,253.2	112,298.3

Note: Values are billions of rupiah at constant 2010 prices.

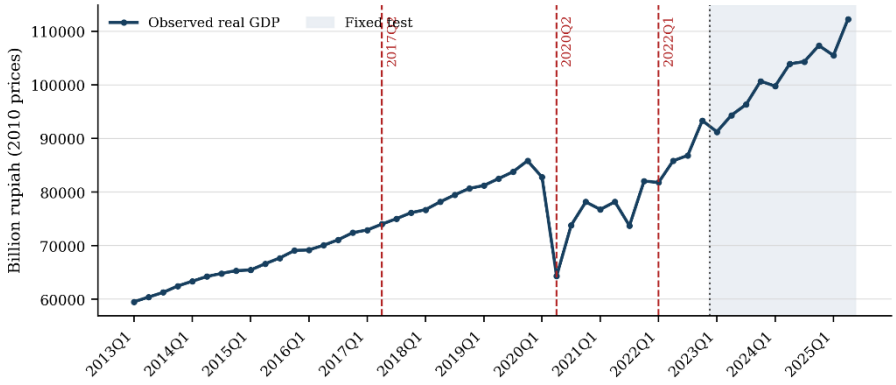

Figure 1. Quarterly real GDP proxy, analytical partition, and selected changepoints

The average log quarter-to-quarter growth was highest in Q4 (3.57%) and lowest in Q1 (-0.56%). Nevertheless, the visually recurring quarterly movement was much smaller than the pandemic disruption. This distinction later explains why a quarterly frequency was retained but the best holdout model required no seasonal AR or MA coefficient.

Table 4. Stationarity-test results for the training sample

Series	ADF statistic	ADF decision	KPSS statistic	KPSS decision
ln(Y)	-1.351	Non-stationary	0.510	Non-stationary
Δln(Y)	-7.556	Stationary	0.110	Stationary
Δ4ln(Y)	-3.376	Stationary	0.118	Stationary
ΔΔ4ln(Y)	-8.138	Stationary	0.177	Stationary

Note: Five-percent critical values were −2.93 for ADF and 0.463 for KPSS. Bold indicates the transformation used to set d=1.

The log level failed both stationarity criteria: the ADF statistic (−1.351) did not cross the left-tail critical value, and the KPSS statistic (0.510) exceeded its critical value. By contrast, the first log difference satisfied both tests (ADF=−7.556; KPSS=0.110). Seasonal and combined differences were also stationary, but applying unnecessary seasonal differencing would consume observations and could amplify noise. The joint evidence therefore supported d=1 and allowed D to be determined through model comparison.

3.2. Changepoints and Structural Interpretation

At the primary penalty 1.5ln(n), the penalized PELT objective (Killick et al., 2012) identified changepoints at 2017Q2, 2020Q2, and 2022Q1. The same three dates were selected for multipliers from 0.5 through 2.0. At 2.5ln(n), only 2020Q2 remained, and penalties of at least 3.0ln(n) selected no break. Thus, 2020Q2 was the most penalty-robust structural change, whereas 2017Q2 and 2022Q1 should be interpreted as moderate, penalty-sensitive regime boundaries.

Table 5. PELT segments at the primary penalty

Segment	n	Mean standardized residual	SD
2013Q1–2017Q1	17	-0.152	0.392
2017Q2–2020Q1	12	0.987	0.355
2020Q2–2021Q4	7	-1.487	1.006
2022Q1–2022Q4	4	0.287	0.632

Note: Residuals were obtained after removing a linear trend and quarterly indicators from ln(Y). Segment boundaries begin at the selected changepoint dates.

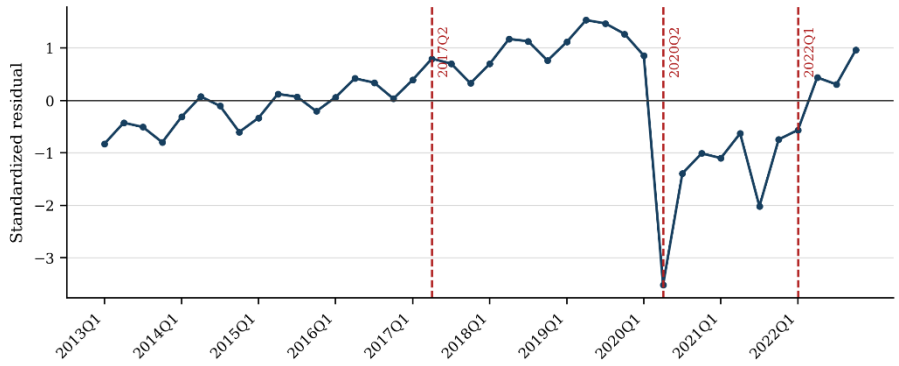

Figure 2. PELT changepoints in detrended and seasonally adjusted log-GDP residuals

The segment mean moved from −0.152 in 2013Q1–2017Q1 to 0.987 in 2017Q2–2020Q1, then fell to −1.487 during 2020Q2–2021Q4 before returning to 0.287 in 2022. These are deviations relative to the fitted deterministic trend and quarter effects, not raw GDP growth rates. Consequently, the 2017Q2 result indicates a statistical level shift in the residual process but does not, by itself, identify an economic cause. The 2020Q2 break is different because it coincides with the largest observed contraction and remains under the most conservative penalty that detects any change.

Economically, Category I output fell from Rp82,788.0 billion in 2020Q1 to Rp64,334.8 billion in 2020Q2, equivalent to -22.29% quarter-to-quarter and -22.01% year-to-year on the ordinary percentage scale. Over the same quarter, Indonesia's aggregate economy contracted by 4.19% quarter-to-quarter and 5.32% year-to-year (BPS, 2020a). The much larger decline in accommodation and food service output is consistent with the greater exposure of contact-intensive and visitor-facing activities.

The 2022Q1 boundary marks a movement out of the negative residual regime, but it should not be treated as the exact start date of recovery. By 2022Q4, the proxy reached Rp93,352.2 billion, 8.7% above its pre-pandemic 2019Q4 level. This timing is consistent with the broader national recovery: BPS reported 5.31% growth in 2022 compared with 3.70% in 2021 (BPS, 2023). The alignment supports an economic interpretation, while the PELT result itself remains associational rather than causal.

3.3. Model Screening, Rolling Validation, and Parameter Meaning

AICc and rolling validation produced different rankings. Within the SARIMA family, SARIMA $(1,1,2)\times(0,0,0)_4$ had the smallest training AICc (−113.918), whereas SARIMA $(0,1,1)\times(0,0,0)_4$ produced the lowest rolling RMSE among the screened SARIMA orders. Likewise, additive Holt–Winters led the ETS family by training AICc, but SES had the smallest rolling RMSE. This difference is expected because AICc measures penalized in-sample fit, while rolling validation directly evaluates repeated forecast errors.

Table 6. Rolling-origin validation accuracy of family representatives

Model	Family	MAE	RMSE	sMAPE	MASE
SES	ETS/Holt–Winters	4,461.1	6,283.9	5.76%	0.854
SARIMA-I $(0,1,1)\times(0,0,0)_4$ + pulse/temp	SARIMA intervention	3,871.4	6,917.4	4.81%	0.742
SARIMA $(0,1,1)\times(0,0,0)_4$	SARIMA	4,919.7	8,054.7	6.57%	0.942
Seasonal-naive	Baseline	7,289.5	8,666.5	9.26%	1.396

Note: Sixteen one-step origins, 2019Q1–2022Q4. Lower values indicate better accuracy. Bold marks the smallest rolling RMSE.

SES achieved the smallest rolling RMSE (Rp6,283.9 billion), while the intervention model had the smallest MAE, sMAPE, and MASE. Because the preregistered family selection criterion was RMSE, SES was the overall rolling-validation winner and the displayed SARIMA-intervention specification was the winner within its own family. This divergence across loss functions shows that model superiority was not uniform: SES avoided a few large errors, whereas the intervention model was closer on the typical origin. For the selected SARIMA representative, the fitted log-difference equation was.

$$\Delta z_t = 0.01074 + \varepsilon_t - 0.39964\varepsilon_{t-1} \quad (22)$$

The drift of 0.01074 corresponds to a baseline quarterly log growth of about 1.07% or 4.39% when compounded over four quarters. The MA (1) coefficient of -0.39964 implies partial correction of the preceding innovation: a temporary positive deviation is followed by an offsetting adjustment, and vice versa. No seasonal AR or MA term was retained. Mathematically, this means that first differencing and short-memory error correction carried more stable predictive information than explicit seasonal dependence in this short, shock-affected quarterly sample.

The intervention coefficients were −0.1669 for the 2020Q2 pulse and −0.0918 for the temporary 2020Q2–2021Q4 indicator. On the multiplicative scale these correspond to conditional level shifts of approximately -15.4% and -8.8%, with a combined 2020Q2 shift of about -22.8%. These values quantify the shock conditional on the specified dynamics; they are not causal treatment effects. Their economic plausibility does not automatically make the intervention model the best forecaster.

3.4. Residual Adequacy and Fixed-Test Performance

Table 7. Residual diagnostics for the family representatives

Model	LB p(4)	LB p(8)	JB p	ARCH-LM p	Inference
SARIMA	0.874	0.817	<0.001	0.998	White noise; non-normal; no ARCH
SARIMA intervention	<0.001	0.001	<0.001	0.052	Autocorrelation remains; non-normal
SES	0.820	0.765	<0.001	0.994	White noise; non-normal; no ARCH

Note: LB=Ljung–Box; JB=Jarque–Bera. The five-percent level was used.

The SARIMA residuals passed the Ljung–Box tests at lags 4 and 8 (p=0.874 and p=0.817), and the ARCH-LM test showed no remaining conditional heteroskedasticity (p=0.998). SES produced a similar white-noise conclusion. All three models rejected residual normality, reflecting the heavy-tailed pandemic shock. The intervention model reduced residual scale but retained significant serial dependence (both Ljung–Box p-values below 0.002), indicating that the pulse and temporary dummy did not fully represent the recovery path. Accordingly, residual autocorrelation was model-specific: it was absent in the selected baseline SARIMA but remained in the intervention specification. For that reason, the intervention model is useful for structural explanation and stress-testing scenarios but weaker as the principal stochastic forecasting model.

Table 8. Fixed ten-quarter test accuracy, 2023Q1–2025Q2

Model	MAE	RMSE	sMAPE	MASE
SARIMA$(0,1,1)\times(0,0,0)_4$	5,301.9	6,131.6	5.23%	1.016
SARIMA-I $(0,1,1)\times(0,0,0)_4$ + pulse/temp	6,207.9	7,056.9	6.17%	1.189
SES	9,932.4	11,549.6	10.09%	1.902
Seasonal-naive	15,251.4	16,491.6	16.14%	2.921

Note: The test set was not used for order or parameter selection. Bold marks the smallest RMSE and sMAPE.

On the untouched fixed test, SARIMA$(0,1,1)\times(0,0,0)_4$ had the lowest RMSE (Rp6,131.6 billion) and sMAPE (5.23%). Its RMSE was 13.1% lower than the intervention model, 46.9% lower than SES, and 62.8% lower than the seasonal-naïve benchmark. The appropriate wording is therefore that SARIMA was the best-performing final holdout model, not that it had won every validation exercise.

The reversal between rolling validation and fixed-test rankings is substantively important. SES performed well when successive one-quarter forecasts spanned both normal and pandemic periods, but after being fitted through 2022Q4 it produced a flat ten-quarter projection of about Rp91,756.1 billion. It therefore failed to follow the continued recovery. SARIMA's positive drift adapted better to the higher post-2022 trajectory, although its MASE of 1.016 shows that its test MAE remained slightly above the average in-sample seasonal-naïve scale. This prevents overstating the result as universally high accuracy.

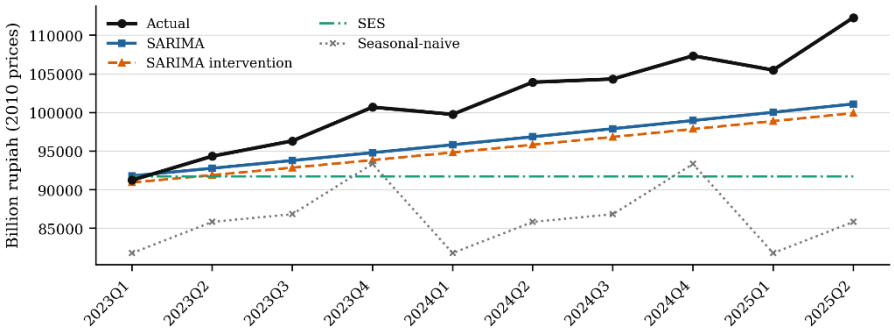

Figure 3. Actual values and fixed-test forecasts, 2023Q1–2025Q2

Table 9. Fixed-test forecasts from the best-performing holdout model

Period	Actual	SARIMA forecast	Error (actual−forecast)	APE
2023Q1	91,253.2	91,804.8	-551.6	0.60%
2023Q2	94,357.7	92,796.0	1,561.7	1.66%
2023Q3	96,336.4	93,798.0	2,538.4	2.63%
2023Q4	100,704.3	94,810.7	5,893.6	5.85%
2024Q1	99,772.7	95,834.5	3,938.2	3.95%
2024Q2	103,937.2	96,869.2	7,068.0	6.80%
2024Q3	104,350.0	97,915.2	6,434.8	6.17%
2024Q4	107,357.9	98,972.4	8,385.5	7.81%
2025Q1	105,511.1	100,041.0	5,470.1	5.18%
2025Q2	112,298.3	101,121.2	11,177.1	9.95%

Note: Values are billions of rupiah at constant 2010 prices. Positive error indicates underprediction.

Forecast errors became increasingly positive after 2023Q1, indicating systematic underprediction as the recovery strengthened. The largest displayed gap occurred in 2025Q2: the actual value was Rp112,298.3 billion compared with a forecast of Rp101,121.2 billion, an underprediction of Rp11,177.1 billion (9.95%). BPS reported that the overall economy grew 4.04% quarter-to-quarter and 5.12% year-to-year in 2025Q2 (BPS, 2025a), while the proxy's log growth was 6.23% and 7.74%, respectively. The stronger sectoral expansion explains why a model estimated only through 2022Q4 lagged the later trajectory. Parameter stability was not formally established, and the increasingly one-sided errors suggest that the estimated drift may not fully represent the evolving post-pandemic recovery regime.

3.5. Mathematical, Economic, and Managerial Implications

3.5.1. Mathematical Implication

The results demonstrate that a parsimonious integrated moving-average process can outperform more elaborate seasonal and intervention specifications when the sample is short and interrupted. Complexity improved training fit but did not guarantee lower rolling or holdout error. PELT and SARIMA answered different questions: PELT located changes in the deterministic-adjusted distribution, whereas SARIMA summarized the conditional stochastic evolution. Their combination is therefore complementary rather than redundant. The failure of residual normality also means that Gaussian prediction intervals would be fragile, even though point forecasts were useful. Consequently, the reported trajectory should be interpreted as a central forecasting baseline rather than a complete representation of forecast uncertainty.

3.5.2. Economic Implication

The 2020Q2 break separates an ordinary growth path from an exceptional contraction in visitor-facing output, while the 2022Q1 boundary marks a statistical normalization phase. The continued rise through 2025Q2 indicates that the post-pandemic path was not merely a return to the old level; it moved above the pre-pandemic maximum. However, because Category I also includes food-service consumption by residents and non-tourism activity, the estimates represent tourism-related production exposure rather than complete tourism direct GDP.

3.5.3. Managerial Implication

Tourism agencies, hotel and restaurant operators, and local planners can use the SARIMA trajectory as a central quarterly baseline for staffing, procurement, and capacity planning, but not as a fixed budget ceiling. The late-horizon underprediction supports maintaining an upside capacity buffer during recovery. Forecasts should be refreshed each quarter, and large standardized residuals should trigger changepoint monitoring and model re-estimation. During future disruptions, the intervention specification remains useful for scenario explanation, while operational decisions should compare several models rather than relying on a single curve.

3.6. Relation to Previous Research and Limitations

The final holdout ranking is consistent with Fahmiyah et al. (2024), who found SARIMA more accurate than Holt–Winters for monthly foreign tourist arrivals to Indonesia. The present evidence extends that comparison to quarterly real accommodation and food-service output and adds rolling validation, a fixed holdout, changepoint detection, an intervention family, and residual diagnostics. The selected orders differ because tourist arrivals and real sectoral value added are different processes and because explicit quarterly seasonality was not stable enough to improve this study's holdout forecasts.

The findings also support the broader argument that interrupted series require model-strategy comparison rather than automatic insertion of a shock dummy (Hyndman & Rostami-Tabar, 2025). Here, the intervention coefficients were economically interpretable, but the intervention model's residual autocorrelation and weaker holdout accuracy prevented it from becoming the principal forecast. Its appropriate role is structural interpretation and scenario stress testing rather than replacement of the better-performing baseline SARIMA forecast. Similarly, tourism recovery research has emphasized scenario adjustment because pandemic paths are not fully represented by historical regularities (Zhang et al., 2021). For management use, the point forecasts should therefore be accompanied by alternative recovery and disruption scenarios.

Five limitations qualify the results. First, 50 quarterly observations constrain the reliable complexity of all models and make pairwise forecast-significance tests underpowered. Second, the final evaluation contained only ten quarters, so ranking uncertainty remains material. Third, no external predictors such as tourist arrivals, occupancy, visitor expenditure, exchange rates, or tourism foreign exchange receipts were included. Fourth, only point forecasts were evaluated; the strongly non-normal residuals make conventional Gaussian intervals inappropriate without bootstrap or other robust alternatives. Fifth, parameter stability was not formally tested, and the systematic late-horizon underprediction indicates that the post-pandemic drift may continue to evolve. These limitations motivate future SARIMAX or dynamic-regression extensions using additional tourism indicators, residual-bootstrap prediction intervals, and rolling parameter-stability assessment while preserving the same leakage-free validation design.

Overall, the combined evidence answers the research objective in three layers. PELT establishes that the series contains statistically meaningful regime changes, particularly in 2020Q2. The forecasting comparison shows that parsimonious SARIMA dynamics generalized best to the fixed post-2022 holdout. Finally, the economic and managerial interpretation converts the mathematical results into a cautious monitoring framework: use SARIMA as the central baseline, retain intervention models for shock scenarios, and update the model whenever new data indicate a structural departure.

4. CONCLUSION

This study examined structural changes and forecast quarterly real GDP for Indonesia's Accommodation and Food Service Activities using 50 observations from 2013Q1 to 2025Q2. The response was treated as a constant-price proxy for tourism-related economic output, not as complete Tourism Direct Gross Domestic Product. After trend and quarterly effects were removed from the log-transformed training series, PELT identified changepoints at 2017Q2, 2020Q2, and 2022Q1 under the primary penalty. The same three dates persisted across penalties from 0.5ln(n) to 2.0ln(n), while 2020Q2 remained the sole break at a stronger penalty. These results indicate several regimes, with the most robust disruption coinciding with the pandemic contraction and the 2022Q1 break marking the subsequent recovery phase. The dates should be interpreted as distributional changes associated with economic events rather than as stand-alone evidence of causality.

The forecasting comparison produced a deliberately nuanced answer to the model-selection objective. Simple exponential smoothing achieved the smallest one-step rolling-validation RMSE (6,283.9), whereas the pulse-plus-temporary-step SARIMA-intervention model was the best specification within the intervention family (RMSE 6,917.4). However, when every specification was frozen and evaluated on the untouched 2023Q1-2025Q2 test period, SARIMA$(0,1,1)\times(0,0,0)_4$ performed best, with MAE of Rp5,301.9 billion, RMSE of Rp6,131.6 billion, sMAPE of 5.23%, and MASE of 1.016. It outperformed the selected SARIMA-intervention, SES, and seasonal-naive models on the final holdout. The selected baseline SARIMA residuals showed no significant autocorrelation at lags 4 and 8 and no ARCH effect, whereas the intervention model retained significant residual autocorrelation. Strong non-normality across the retained models limits Gaussian uncertainty statements. Forecast errors also became increasingly positive during the later recovery, indicating that the fitted model tended to underpredict the accelerating output path.

Mathematically, the results show that changepoint evidence, intervention fit, and forecast accuracy are related but distinct criteria. Detecting a major interruption does not guarantee that an intervention specification will generalize better than a parsimonious stochastic model, and a rolling-validation winner need not remain superior over a longer fixed horizon. Economically, the detected regimes reflect contraction and recovery in a sector closely connected with mobility, visitor expenditure, and household food-service demand. For management, the selected SARIMA model can serve as a transparent baseline for quarterly capacity, staffing, procurement, and policy monitoring, while intervention scenarios should be retained for stress testing rather

than imposed automatically. The model should be re-estimated as each new quarter becomes available, and persistent one-sided errors should trigger renewed changepoint assessment or model revision.

The findings are subject to several limitations. The Category I GDP series contains non-tourism activity and therefore cannot isolate the full direct contribution of tourism. The sample contains only 50 quarterly observations, the models are primarily univariate, and the final comparison evaluates point forecasts rather than robust prediction intervals. The extraordinary pandemic shock also produced non-normal residuals, while formal parameter stability was not established and systematic late-horizon underprediction suggests that the post-pandemic drift may continue to evolve. Future studies should extend the series and evaluate dynamic-regression or SARIMAX models using international visitor arrivals, tourism foreign exchange receipts, visitor expenditure, exchange rates, and other contemporaneous indicators. Subnational or sector-specific panels, robust or bootstrap prediction intervals, and regime-switching or time-varying-parameter models would provide useful extensions. Overall, the study contributes a reproducible, leakage-free PELT-SARIMA-intervention workflow that separates structural diagnosis from predictive verification and translates both into cautious economic and managerial guidance.

REFERENCES

Badan Pusat Statistik. (2017). Produk domestik bruto triwulanan 2013–2017. https://www.bps.go.id/id/publication/2017/10/02/a169618dd6a187b5029ab668/produk-domestik-bruto-triwulanan-2013-2017.html

Badan Pusat Statistik. (2019). PDB Indonesia triwulanan 2015–2019. https://www.bps.go.id/id/publication/2019/10/07/4923ba3ffd04cd25e83dcd97/pdb-indonesia-triwulanan-2015-2019.html

Badan Pusat Statistik. (2020a, August 5). Ekonomi Indonesia triwulan II 2020 turun 5,32 persen. https://www.bps.go.id/id/pressrelease/2020/08/05/1737/ekonomi-indonesia-triwulan-ii-2020-turun-5-32-persen.html

Badan Pusat Statistik. (2020b). PDB Indonesia triwulanan 2016–2020. https://www.bps.go.id/id/publication/2020/10/16/54be7f82b7d3aa22f5e2c144/pdb-indonesia-triwulanan-2016-2020.html

Badan Pusat Statistik. (2023, February 6). Ekonomi Indonesia tahun 2022 tumbuh 5,31 persen. https://www.bps.go.id/id/pressrelease/2023/02/06/1997/ekonomi-indonesia-tahun-2022-tumbuh-5-31-persen.html

Badan Pusat Statistik. (2024). Produk domestik bruto Indonesia triwulanan 2020–2024. https://www.bps.go.id/id/publication/2024/10/09/7290b829d2eaa972e4968d19/produk-domestik-bruto-indonesia-triwulanan-2020-2024.html

Badan Pusat Statistik. (2025a, August 5). Ekonomi Indonesia triwulan II-2025 tumbuh 4,04 persen (q-to-q); 5,12 persen (y-on-y); semester I-2025 tumbuh 4,99 persen (c-to-c). https://www.bps.go.id/id/pressrelease/2025/08/05/2455/ekonomi-indonesia-triwulan-ii-2025-tumbuh-4-04-persen--q-to-q---5-12-persen--y-on-y---semester-i-2025-tumbuh-4-99-persen--c-to-c--.html

Badan Pusat Statistik. (2025b). Produk domestik bruto Indonesia triwulanan 2021–2025. https://www.bps.go.id/id/publication/2025/10/16/04bf932a1a65d15545e6de15/produk-domestik-bruto-indonesia-triwulanan-2021-2025.html

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed.). Wiley.

Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical Association, 70(349), 70–79. https://doi.org/10.1080/01621459.1975.10480264

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366a), 427–431. https://doi.org/10.1080/01621459.1979.10482531

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), 987–1007. https://doi.org/10.2307/1912773

Fahmiyah, I., Andini, L. S., & Ghani, M. (2024). Forecasting the number of foreign tourism visits to Indonesia using seasonal autoregressive integrated moving average (SARIMA) and Holt–Winters approach. In T. Amrillah et al. (Eds.), Proceedings of the International Conference on Advanced Technology and Multidiscipline (ICATAM 2024) (pp. 354–371). Atlantis Press. https://doi.org/10.2991/978-94-6463-566-9_23

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. Biometrika, 76(2), 297–307. https://doi.org/10.1093/biomet/76.2.297

Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice (3rd ed.). OTexts. https://otexts.com/fpp3/

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting, 18(3), 439–454. https://doi.org/10.1016/S0169-2070(01)00110-8

Hyndman, R. J., & Rostami-Tabar, B. (2025). Forecasting interrupted time series. Journal of the Operational Research Society, 76(4), 790–803. https://doi.org/10.1080/01605682.2024.2395315

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics Letters, 6(3), 255–259. https://doi.org/10.1016/0165-1765(80)90024-5

Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. Journal of the American Statistical Association, 107(500), 1590–1598. https://doi.org/10.1080/01621459.2012.737745

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics, 54(1–3), 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297–303. https://doi.org/10.1093/biomet/65.2.297

Rianda, F., & Usman, H. (2023). Forecasting tourism demand during the COVID-19 pandemic: ARIMAX and intervention modelling approaches. BAREKENG: Jurnal Ilmu Matematika dan Terapan, 17(1), 285–294. https://doi.org/10.30598/barekengvol17iss1pp0285-0294

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. International Journal of Forecasting, 16(4), 437–450. https://doi.org/10.1016/S0169-2070(00)00065-0

UN Tourism. (2025, January 21). International tourism recovers pre-pandemic levels in 2024. https://www.unwto.org/news/international-tourism-recovers-pre-pandemic-levels-in-2024

Zhang, H., Song, H., Wen, L., & Liu, C. (2021). Forecasting tourism recovery amid COVID-19. Annals of Tourism Research, 87, Article 103149. https://doi.org/10.1016/j.annals.2021.103149