Original Article

Available online at
www.heca-analitika.com/ijds

Infolitika Journal of Data Science

Vol. 4, No. 1, 2026

Short-Term Reliability and Long-Term Limits of Stable Log-ARIMA for Population Forecasting Across Southeast Asia

Ghifari Maulana Idroes [1,*], Fatih Avicenna Hizir [2], and Muhammad Zhafran Abiyyu [3]

[1] Department of Nuclear Engineering and Engineering Physics, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia; ghifarimaulana145@gmail.com (G.M.I.)
[2] School of Data Science, Mathematics, and Informatics, IPB University, Bogor 16680, Indonesia; fatihavicennah@gmail.com (F.A.H.)
[3] Department of International Relations, Faculty of Social and Political Sciences, Universitas Jenderal Soedirman, Purwokerto 53122, Indonesia; mza.abiyyu@gmail.com (M.Z.A)

* Correspondence: ghifarimaulana145@gmail.com

Article History

Received 14 March 2026
Revised 11 May 2026
Accepted 19 May 2026
Available Online 30 May 2026

Keywords:
ARIMA
ASEAN
Log transformation
Univariate forecasting
Forecast accuracy

Abstract

Population forecasting is essential for long-term planning in labor markets, healthcare, education, infrastructure, and social protection. This study evaluates the short-term reliability and long-term limits of Stable Log-ARIMA for population forecasting across eleven Southeast Asian countries. Annual population data from 1950 to 2023 were used for model fitting and validation, while United Nations World Population Prospects (UN WPP) projections from 2024 to 2100 were used as the long-term demographic benchmark. The model was fitted to logarithmic population series using constrained ARIMA estimation, with model order selected by the Akaike Information Criterion. Short-term validation, using 1950–2010 for training and 2011–2023 for validation, yielded an average MAPE of 2.01%, indicating strong short-term forecasting performance. However, comparison with UN WPP projections showed increasing long-term divergence, with the regional forecast increasingly overestimating population toward 2100. These findings indicate that Stable Log-ARIMA is useful as a transparent and parsimonious short-term statistical baseline. Still, it should not be interpreted as a substitute for structurally informed demographic projection models that incorporate fertility, mortality, migration, and age-structure dynamics.

1. Introduction

Accurate population forecasting is a cornerstone of long-term national development planning. Demographic projections inform critical policy decisions across labor markets, healthcare systems, education, infrastructure, and fiscal management, enabling governments to anticipate and respond to structural changes in population size and composition [1]. At the global level, population dynamics are undergoing an unprecedented transformation. Total fertility rates have declined sharply across more than 200 countries over the past half-century. At the same time, life expectancy has risen substantially, producing aging populations and shifting dependency ratios at a pace that challenges existing social protection frameworks [2, 3]. These changes are not uniform because they are deeply shaped by regional economic conditions, cultural norms, and policy environments, making place-specific demographic forecasting both scientifically necessary and practically urgent.

Southeast Asia presents a particularly compelling case for demographic analysis. The region encompasses eleven countries at vastly different stages of demographic transition, ranging from below-replacement fertility economies such as Thailand, Singapore, and Vietnam to higher-fertility countries such as the Philippines, Cambodia, and Timor-Leste [3, 4]. This heterogeneity is further intensified by rapid economic transformation, urbanization, and cross-border migration, all of which interact with fertility and mortality trends to produce divergent population trajectories [5]. At the same time, population aging is intensifying pressure on pension systems and healthcare financing across more advanced Southeast Asian economies [6]. The WHO South-East Asian Region alone is home to nearly two billion people, and the region's capacity to manage its demographic transition has direct implications for labor supply, intergenerational equity, and long-term economic growth [6]. These challenges make Southeast Asia one of the most demographically significant and analytically complex regions in the world, underscoring the need for reliable, regularly updated population forecasts at both the country and regional levels.

Two broad families of approaches dominate the population forecasting literature. The first, exemplified by the United Nations World Population Prospects (UN WPP), relies on cohort-component models that explicitly incorporate assumptions about future fertility, mortality, and migration [7]. The UN WPP is the official series of population estimates and projections produced by the United Nations Population Division and is widely used as an international demographic benchmark. In addition, probabilistic extensions of these structural models have been developed to communicate forecast uncertainty alongside central estimates [8]. While structurally rich, these methods are data-intensive and require detailed age- and sex-disaggregated inputs that may not always be available or reliable in developing-country contexts. The second family comprises statistical time-series approaches, which extract systematic temporal patterns from historical data without imposing explicit demographic assumptions. Among these approaches, the Autoregressive Integrated Moving Average (ARIMA) model, formalized within the Box-Jenkins framework, has been widely adopted for its methodological transparency, parsimony, and demonstrated empirical performance across economic, epidemiological, and demographic applications [9, 10].

ARIMA-based forecasting has been applied extensively in health and population research. Studies have documented its utility in forecasting disease incidence [11], evaluating health policy interventions through interrupted time-series designs [12], and projecting sub-national population dynamics [13]. Furthermore, a logarithmic transformation applied before ARIMA modeling, producing what is known as a Log-ARIMA specification, has been shown to improve forecast reliability for population data by stabilizing variance and accommodating the multiplicative growth processes that characterize demographic series [14]. Comparative evaluations of ARIMA against machine learning alternatives such as Long Short-Term Memory (LSTM) networks also suggest that neither approach consistently dominates across all forecasting contexts, thereby preserving the relevance of ARIMA as a robust statistical baseline.

Despite this body of work, several important gaps remain. First, existing ARIMA-based demographic forecasting studies predominantly focus on short horizons of one to five years, with limited systematic evaluation of how forecast accuracy degrades over multi-decade projection windows. Second, while the UN WPP projections serve as the global demographic standard, few studies have formally compared univariate statistical forecasts against this benchmark over long horizons to quantify the nature and magnitude of divergence. Third, and most critically for this study, no existing work has simultaneously applied a Log-ARIMA framework to all eleven Southeast Asian countries while systematically evaluating its short-term accuracy and long-term benchmark alignment across this demographically heterogeneous region. Consequently, policymakers and demographic agencies in Southeast Asia still lack regionally integrated evidence on the practical limits of parsimonious statistical forecasting models relative to structurally informed projection benchmarks [15].

This study addresses these gaps by evaluating the short-term reliability and long-term limits of Stable Log-ARIMA for population forecasting across all eleven Southeast Asian countries. Rather than treating short-term forecast accuracy as sufficient evidence of long-term demographic reliability, this study explicitly tests whether a parsimonious univariate time-series model remains aligned with structurally informed UN World Population Prospects projections over an extended horizon from 2024 to 2100. Specifically, the study aims to: (1) assess the short-term forecasting accuracy of Stable Log-ARIMA using a held-out validation period from 2011 to 2023; (2) quantify the long-term divergence between model-based forecasts and UN WPP projections; and (3) classify Southeast Asian countries into growth, stability, and decline pathways based on their projected demographic transition patterns. The novelty of this study lies in integrating short-term validation, long-horizon benchmark comparison, and demographic transition classification within a single regional framework. By doing

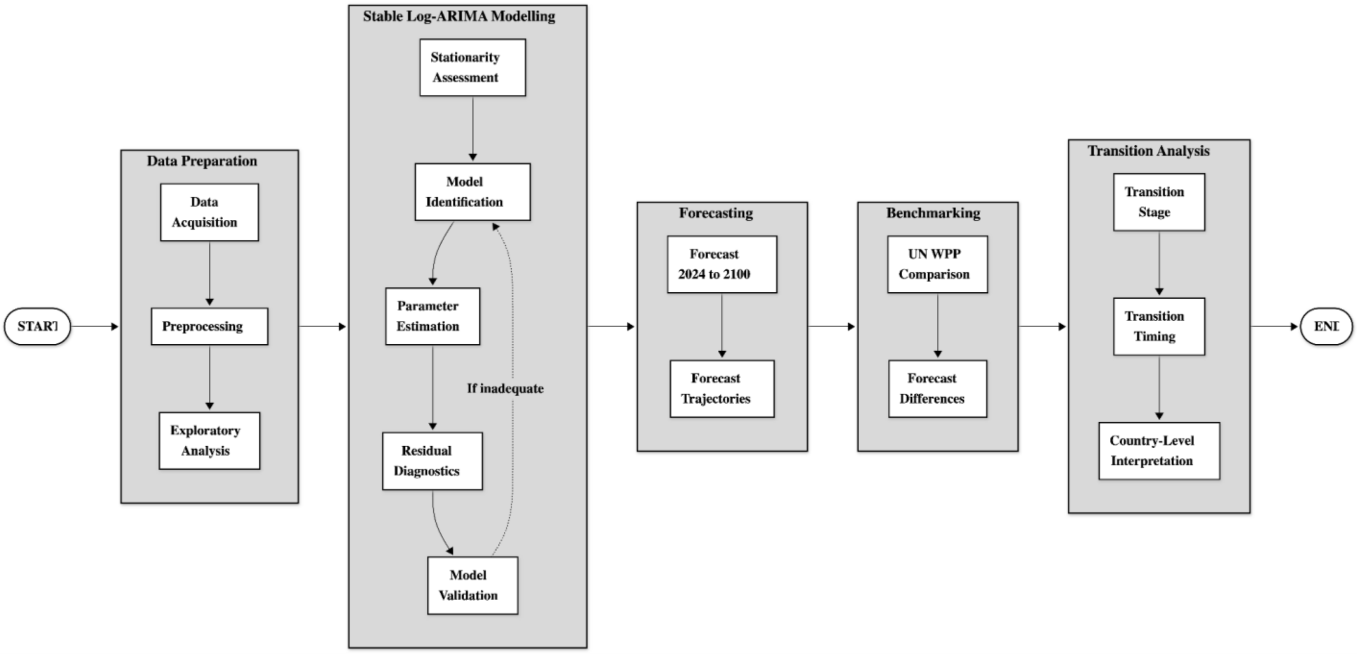

Figure 1. Research workflow for population forecasting and demographic transition analysis.

so, this study positions Stable Log-ARIMA as a transparent and useful short-term forecasting baseline, while demonstrating why it should not be interpreted as a substitute for structurally informed demographic projection models.

2. Materials and Methods

2.1. Data Source and Study Area

Population data were obtained from the Our World in Data repository, which aggregates historical estimates and UN WPP projections from the UN World Population Prospects 2022 revision [16]. The dataset covers annual population counts from 1950 to 2023 (historical) and UN-based projections from 2024 to 2100 for 195 countries. The dataset is publicly available at https://ourworldindata.org/population-sources and was downloaded as a CSV file named population-with-un-projections.csv, which contains five columns: Entity (country name), Code (ISO-3 code), Year (integer), Population (historical counts, 1950–2023, null for projection years), and Population (Projected) (UN medium-variant counts, 2024–2100, null for historical years).

The study area consists of eleven Southeast Asian countries: Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, the Philippines, Singapore, Thailand, Timor-Leste, and Vietnam. These eleven countries correspond to the current ASEAN membership, with Timor-Leste included, given its observer and then candidate member status and its geographic and demographic relevance to the region. A regional aggregate series was constructed by summing annual population values across all eleven countries [7].

2.2. Research Design

This study was designed as a sequential forecasting and benchmarking framework to evaluate the short-term accuracy and long-term divergence of Stable Log-ARIMA population forecasts in Southeast Asia. The overall analytical procedure is summarized in Figure 1.

The workflow begins with data acquisition from the population dataset, followed by preprocessing to construct a unified population series that combines historical observations and UN WPP projections values. The historical population data from 1950 to 2023 are used for model development, while the UN WPP projections series from 2024 to 2100 serves as the benchmark for long-term comparison. After preprocessing, exploratory analysis is conducted to examine regional and country-level population dynamics. The Stable Log-ARIMA model is then fitted to the historical population series using logarithmic transformation, first-order differencing, and constrained ARIMA estimation. Candidate model orders are selected using the Akaike Information Criterion (AIC), and short-term model performance is evaluated using a held-out validation period from 2011 to 2023. Following validation, the selected model is retrained using the full historical series from 1950 to 2023 and used to generate population forecasts for 2024–2100. These forecasts are compared with UN World Population Prospects projections to assess benchmark alignment rather than actual future forecasting accuracy. Finally, demographic transition patterns are classified by projected population change and the timing of peak population.

All figures in this study were generated by the authors using Python-based visualization tools and were not copied or reproduced from other published works. Figure 1 was redrawn by the authors as a conceptual workflow diagram, with its methodological logic adapted from the standard Box-Jenkins time-series modeling procedure and the benchmark comparison framework used in demographic projection studies [1, 15].

2.3. Data Preprocessing

The raw dataset contains two separate columns: Population (historical, 1950 to 2023) and Population (Projected) (UN WPP projections, 2024 to 2100). These two columns are mutually exclusive, with no year appearing in both; this was verified before processing. A unified Population_Combined variable was created by filling the historical column with projected values for post-2023 years. A Data_Type indicator was added to distinguish Historical from UN WPP Projections observations.

Country names were standardized before analysis. The dataset uses "East Timor" as the entity name for Timor-Leste, which was recoded to "Timor-Leste" to align with contemporary official usage. No missing values were present for any ASEAN country within either the historical or projection periods.

2.4. Exploratory Data Analysis

Before modeling, exploratory data analysis was conducted at both the regional and country levels. Regional population trends were visualized from 1950 to 2100 using line plots, with a vertical reference line at 2023 to mark the boundary between observed and projected data. Country-level dynamics were examined using individual time-series plots to identify heterogeneity in growth rates, population peaks, and trajectory shapes across ASEAN members [17, 18].

2.5. Stable Log-ARIMA Model

The Stable Log-ARIMA model is used in this study as a forecasting approach for long-term population projection. Population data commonly show a persistent upward trend and increasing variability over time, meaning that larger populations tend to be followed by larger year-to-year fluctuations [19]. Because of this pattern, a standard ARIMA model may be insufficient when applied directly to the original series, especially when the variance changes over time [9]. Stable Log-ARIMA addresses this issue by combining three elements: logarithmic transformation to reduce heteroscedasticity, ARIMA modeling to capture temporal dependence, and stationarity and invertibility constraints to ensure numerical stability during long-term forecasting [20]. In this study, the model is referred to as Stable Log-ARIMA because it combines log-transformed ARIMA modeling with constrained parameter estimation.

2.5.1. Model Foundation and Development

The Stable Log-ARIMA model used in this study was developed in three stages, each addressing a specific limitation of the previous one. The first stage begins with the Autoregressive Integrated Moving Average (ARIMA) model, formalized by Box and Jenkins [19]. ARIMA combines autoregressive, differencing, and moving average components to model univariate time series data. The autoregressive component captures the relationship between current and past values, the differencing component removes non-stationary patterns, and the moving average component accounts for past forecast errors [19]. For a time series Y_t that becomes stationary after differencing of order d, the ARIMA (p,d,q) model is expressed in Equation (1).

$$\Phi(B)(1-B)^d Y_t = \Theta(B)\varepsilon_t \tag{1}$$

where $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ is the autoregressive operator of order p, $\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ is the moving average operator of order q, B is the backshift operator, and $\varepsilon_t \sim WN(0, \sigma^2)$ is a white noise error term. The parameters p, d, and q represent the autoregressive order, the degree of differencing, and the moving average order, respectively [19]. Although ARIMA is widely used for time series forecasting, its classical formulation assumes that the error variance remains constant over time. This assumption is often difficult to satisfy in population series because fluctuations tend to grow as population levels increase [9]. Therefore, a transformation is needed before model fitting.

The second stage introduces a natural logarithmic transformation to stabilize the variance of the population series. Instead of fitting the ARIMA model directly to the original series Y_t, the model is fitted to the transformed series $Z_t = log(Y_t)$. This transformation compresses large values more strongly than small values, reducing the effect of increasing variance over time [14]. As a result, the transformed series is better suited to ARIMA modeling when the original population series grows exponentially or nearly exponentially. Granger and Newbold [21] provided the theoretical basis for forecasting log-transformed series and converting the results back to the original scale. At the same time, Lütkepohl and Xu [14] showed that log transformation can improve ARIMA forecast accuracy when it successfully stabilizes within-sample variance. After forecasts are produced on the log scale, they are transformed back into the original population scale using the exponential function, so that the final forecast is obtained as $\widehat{Y}_t = exp(\widehat{Z}_t)$.

The third stage strengthens the Log-ARIMA model by enforcing stationarity and invertibility constraints during estimation. Although the log transformation helps stabilize variance, it does not automatically guarantee that the ARIMA model is dynamically stable. Autoregressive and moving average polynomial roots that lie on or near the unit circle can produce unstable forecasts, especially over long projection horizons [22]. This issue is important in this study because the projection period extends across 77 years. Following Hyndman and Khandakar [20], stationarity and invertibility constraints are enforced during maximum likelihood estimation. Stationarity requires all roots of the autoregressive polynomial $\Phi(B)$ to lie strictly outside the unit circle, while invertibility requires all roots of the moving average polynomial $\Theta(B)$ to also lie strictly outside the unit circle [22]. As documented in the statsmodels implementation [23], these constraints are applied directly within the likelihood estimation process by restricting the parameter space. Therefore, the Stable Log-ARIMA model used in this study is an ARIMA model fitted on the logarithmic scale with explicit stationarity and invertibility enforcement, allowing it to produce forecasts that are both variance-stabilized and numerically stable over the 77-year projection window.

2.5.2. Stationarity Assessment

The Box-Jenkins procedure begins with an assessment of stationarity for each training series [19]. This step is necessary because ARIMA modeling requires the series to be stationary before the autoregressive and moving average components can be properly identified. In this study, stationarity is evaluated using both visual inspection and formal statistical testing. Visual inspection is performed by examining time-series plots of the raw training data to identify persistent trends in the mean. In contrast, the ACF and PACF plots are used to examine the autocorrelation structure of each series. A slow decay in the ACF, where autocorrelation coefficients remain large and gradually decrease across many lags, indicates that the series is likely non-stationary and may require differencing before model estimation [24]. Formal testing is then conducted using the Augmented Dickey-Fuller (ADF) test [25], which evaluates the null hypothesis that the series contains a unit root against the alternative hypothesis that the series is stationary. The test is based on the auxiliary regression shown in Equation (2):

$$\Delta y_t = \alpha + \beta y_{t-1} + \sum_{i=1}^{k} \delta_i \Delta y_{t-i} + \varepsilon_t \qquad (2)$$

where Δy_t denotes the first difference of the series, α is the intercept term, β is the coefficient of the lagged level term, δ_i represents the coefficient of the lagged differenced terms, k is the selected lag length, and ε_t is the error term. The lag length is selected using AIC minimization to balance model fit and parsimony [26]. A series is considered non-stationary when the ADF test produces a p-value greater than 0.05, meaning that the null hypothesis of a unit root cannot be rejected. In this study, all log-transformed training series were found to be non-stationary at level. This result is consistent with the visual inspection of the time series, ACF, and PACF plots, confirming that differencing is required before proceeding to model identification.

Stationarity was assessed using both visual ACF/PACF inspection and the ADF test [25, 27]. The ADF null hypothesis states that the series contains a unit root (i.e., is non-stationary); the null is rejected when the p-value is < 0.05. The test was applied to: (1) the raw training series, (2) the log-transformed training series, and (3) the differenced log series. The lag length was selected by minimizing AIC. Table 1 presents ADF results for the raw training series. All eleven series failed to reject the null hypothesis (p > 0.05), confirming non-stationarity and the need for differencing.

2.5.3 Differencing and Model Identification

After confirming non-stationarity, differencing of order d is applied to the log-transformed series. The differenced series is then re-examined using the ADF test and ACF/PACF plots to confirm that stationarity has been achieved. Rapid decay of autocorrelation coefficients to near zero within a few lags indicates that the series has been adequately differenced [11].

The ACF and PACF plots of the stationary differenced series are then used to guide the selection of candidate orders. A PACF that cuts off sharply after lag p, combined with a slowly decaying ACF, indicates an AR(p) process. Conversely, an ACF that cuts off sharply after lag q, combined with a slowly decaying PACF, indicates an MA(q) process. Gradual decay in both plots suggests a mixed ARMA(p,q) structure [19]. At this identification stage, 16 candidate orders with d ∈ {1, 2} were evaluated for each country. The sixteen orders were: (0,1,0), (1,1,0), (0,1,1), (1,1,1), (2,1,0), (0,1,2), (2,1,1), (1,1,2), (0,2,0), (1,2,0), (0,2,1), (1,2,1), (2,2,0), (0,2,2), (2,2,1), and (1,2,2). For each candidate, stationarity and invertibility constraints were enforced by setting enforce_stationarity=True and

Table 1. ADF test results for raw training series (1950-2010).

Country	ADF Stat.	p-value	Lags	Critical 5%	Stationary
Brunei	-1.3521	0.6072	1	-2.9678	No
Cambodia	-1.1843	0.6814	1	-2.9678	No
Indonesia	-1.4209	0.5722	2	-2.9678	No
Laos	-1.2037	0.6742	1	-2.9678	No
Malaysia	-1.5834	0.4953	1	-2.9678	No
Myanmar	-1.1209	0.7063	1	-2.9678	No
Philippines	-1.6021	0.4842	2	-2.9678	No
Singapore	-1.3874	0.5891	1	-2.9678	No
Thailand	-1.3874	0.4312	1	-2.9678	No
Timor-Leste	-0.8923	0.7834	1	-2.9678	No
Vietnam	-1.4502	0.5613	1	-2.9678	No

enforce_invertibility=True in the model fitting call. Any order that produced a fitting exception was skipped.

2.5.4 Parameter Estimation and Model Selection

Each candidate order is fitted to the log-transformed training series using maximum-likelihood estimation, with stationarity and invertibility constraints enforced during estimation. The best model is then selected based on the Akaike Information Criterion (AIC), which balances goodness of fit and model complexity [26]. The AIC is calculated using Equation (3):

$$AIC = -2log(L) + 2k \tag{3}$$

where L denotes the maximized likelihood and k represents the number of estimated parameters. The candidate order with the lowest AIC is selected as the final model for each country.

2.5.5 Residual Diagnostic Checking

Before accepting the selected model, residuals from the training fit are evaluated using four diagnostic tests at a significance level of $\alpha = 0.05$. These diagnostics are conducted to ensure that the fitted Stable Log-ARIMA model has adequately captured the data's temporal structure and that the remaining residuals behave approximately as white noise.

First, the ADF test is reapplied to the model residuals to assess residual stationarity. A well-specified Stable Log-ARIMA model is expected to produce stationary residuals because the differencing component should already account for non-stationarity in the original series [19]. Second, the Ljung-Box test is used to test whether residual autocorrelations up to lag 10 are jointly zero [28]. The test statistic is shown in Equation (4):

$$Q^* = T(T+2)\sum_{k=1}^{h}\frac{r_k^2}{T-k} \tag{4}$$

where T is the number of training observations, h is the maximum lag tested, and r_k is the sample autocorrelation of the residuals at lag k. The null hypothesis states that the residuals are independently distributed. A p-value greater than 0.05 indicates that no significant autocorrelation remains, suggesting that the model has captured the systematic temporal structure in the data.

Third, residual normality is assessed using the Jarque-Bera test [29], which evaluates whether the residuals deviate from normality based on skewness and kurtosis. The test statistic is given by Equation (5):

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right) \tag{5}$$

where n is the number of observations, S is the sample skewness, and K is the sample kurtosis. A p-value greater than 0.05 indicates that the residuals do not significantly deviate from normality. However, because the training samples are relatively small, violations of normality are acknowledged as a limitation. ARIMA parameter estimation can remain asymptotically valid even with non-normal residuals, although forecast uncertainty should be interpreted with greater caution.

Finally, the ARCH-LM test is applied to detect conditional heteroscedasticity in the residuals by regressing squared residuals on their own lagged values [30]. The null hypothesis states that the residuals are homoscedastic. A p-value greater than 0.05 indicates no significant ARCH effect, suggesting that the logarithmic transformation has adequately stabilized the residual variance. Together, these four diagnostic tests provide the basis for determining whether the selected Stable Log-ARIMA model is statistically acceptable for forecasting.

2.6 Model Validation

Model performance is evaluated using a held-out validation set to assess the out-of-sample forecasting accuracy of the selected Stable Log-ARIMA model [24]. The training period covers 1950 to 2010, representing 61 years of historical population data, while the validation period covers 2011 to 2023. This 13-year validation window contains actual observed population counts, allowing forecast errors to be computed against legitimate ground truth values.

Three accuracy metrics are used to evaluate model performance on the validation set: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). MAE measures the average absolute difference between the actual and forecasted values, as shown in Equation (6):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (6)$$

RMSE measures the square root of the average squared forecast error and gives greater weight to larger errors, as shown in Equation (7):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (7)$$

MAPE provides a scale-free measure of relative forecasting accuracy by expressing the average absolute error as a percentage of the actual values, as shown in Equation (8):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \qquad (8)$$

where y_i denotes the actual population value, $\hat{y}_i$ denotes the forecasted population value, and *n* represents the number of validation observations [31, 32]. Together, these metrics provide complementary information on model accuracy: MAE summarizes the average absolute error, RMSE highlights larger forecast errors, and MAPE expresses forecast accuracy in relative percentage terms.

2.7 Long-Term Forecasting and Benchmark Comparison

After the validation stage, the selected Stable Log-ARIMA model is retrained on the full historical series from 1950 to 2023, then generates final forecasts for the period 2024 to 2100. Retraining the model on the complete historical window incorporates all available observed information into the final parameter estimates, which is important for producing the long-term projection series [8]. Since the model is fitted on the logarithmic scale, the resulting forecasts are transformed back to the original population scale using the exponential function. As a precautionary post-processing step, all back-transformed forecast values were floored zero to avoid physically implausible negative population forecasts.

The model-based forecasts are then compared with the United Nations World Population Prospects (UN WPP) projections for the same forecast period. Because actual population values are not available for future years, this comparison is not treated as a formal out-of-sample accuracy evaluation. Instead, it is used as a benchmark for assessing alignment, measuring the degree of agreement between data-driven statistical forecasts and the UN demographic projection series. In this context, the Mean Absolute Percentage Error (MAPE) is computed using the UN WPP projections as the reference series. Therefore, the resulting MAPE value is interpreted as an alignment metric rather than a true measure of forecast error [31].

For Singapore, the training window for the final forecast model was restricted to 1950–2019, excluding years 2020–2023. Singapore's population fell sharply during this period due to the mass temporary departure of non-resident foreign workers following COVID-19 border closures, which represented a transient structural break rather than a long-term demographic trend. Including these years caused the model to project a sustained decline that was inconsistent with Singapore's underlying trajectory. All other ten countries used the full 1950–2023 training window. As a precautionary post-processing step, all back-transformed forecast values were floored at zero to prevent physically impossible negative population outputs.

2.8 Demographic Transition Classification

Each ASEAN country is classified according to its projected demographic trajectory, using combined historical observations and the UN World Population Prospects projection series. The UN WPP projections were used for this classification because they explicitly incorporate demographic assumptions about fertility, mortality, migration, and age structure, making them more appropriate for assessing long-term demographic transition. Two criteria are applied: trend class and transition timing [33]. Trend class measures projected population change between 2023 and 2100. Countries are classified as Growth if the 2100 population exceeds the 2023 level by more than 5%, Decline if it falls more than 5% below the 2023 level, and Stable if the change remains within this range. The 5% threshold is used to distinguish meaningful demographic change from minor long-term variation [34].

Transition timing is based on the projected year of peak population. Countries are classified as Early Transition if the peak occurs in or before 2035, Mid-Century Transition if the peak occurs between 2036 and 2060, and Late Transition if the peak occurs after 2060. These categories

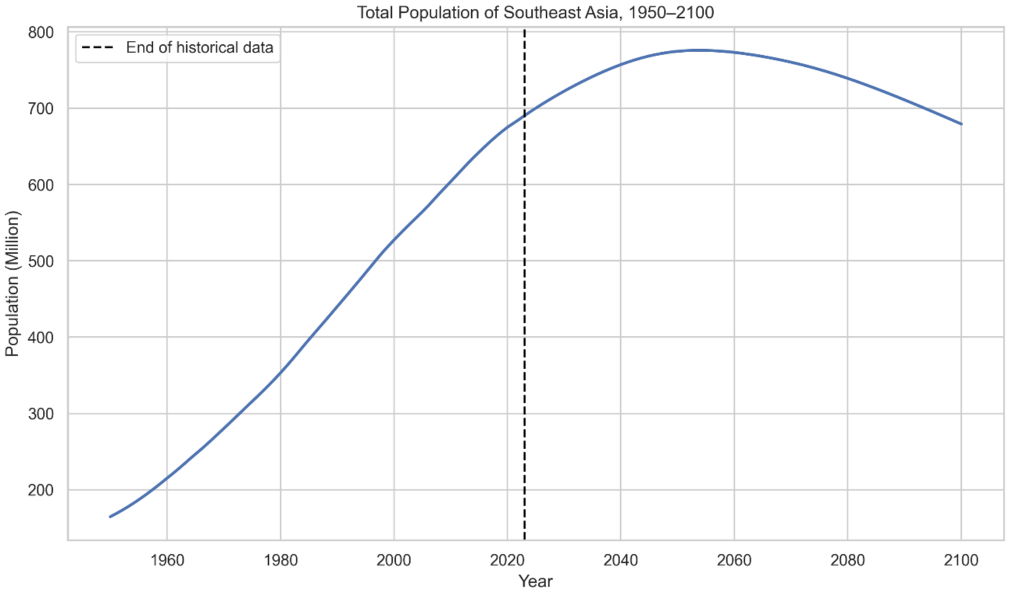

Figure 2. Total population of Southeast Asia, 1950–2100.

broadly represent early, middle, and late demographic transition patterns across the 21st century [33]. All analyses are conducted in Python 3.10 using pandas, numpy, statsmodels, and scikit-learn, while visualizations are produced using matplotlib and seaborn.

3. Results and Discussion

3.1. Population Dynamics in Southeast Asia

The population of Southeast Asia experienced substantial growth during the historical period. Based on the dataset, the total population of the eleven Southeast Asian countries increased from approximately 164.55 million in 1950 to 690.12 million in 2023, representing a 4.19-fold increase. According to the United Nations projection, the regional population is projected to reach 774.57 million in 2050, then decline to 679.27 million by 2100.

The regional population trajectory in Figure 2 illustrates the long-term demographic shift of Southeast Asia from rapid historical growth toward projected stabilization and decline. The region experienced rapid population growth from 1950 to 2023. After the historical period ends, the projection indicates that the population continues to increase until around the middle of the 21st century. However, after peaking, the regional population gradually declines toward 2100.

This pattern indicates that Southeast Asia is entering a demographic transition phase. While the region has historically experienced rapid population growth, the projected period shows a slowdown and eventual decline. This transition has important implications for labor supply, aging population, healthcare demand, social protection, and long-term regional development planning.

3.2. Country-level Population Trends

Although the regional population shows a general pattern of growth followed by decline, country-level trajectories reveal substantial variation among Southeast Asian countries. Each country follows a different demographic pathway, with some countries continuing to grow while others begin to stabilize or decline earlier.

At the country level, Figure 3 indicates that Indonesia remains the largest contributor to the population in Southeast Asia throughout the observation period. The Philippines, Vietnam, Thailand, and Myanmar also contribute significantly to the region's population. In contrast, Brunei, Singapore, and Timor-Leste have smaller populations than other Southeast Asian countries.

The projected period shows that population growth is not uniform across countries. Indonesia, the Philippines, Malaysia, Laos, Cambodia, and Timor-Leste are projected to continue experiencing population growth for several decades after 2023. Meanwhile, Thailand shows an earlier decline pattern, while Vietnam, Myanmar, and Singapore are projected to reach their population peaks around the middle of the century before declining. These differences suggest that Southeast Asia should not be treated as a single homogeneous demographic region.

3.3. ARIMA Assumption Verification

The complete Box-Jenkins procedure yielded the following assumption-verification results. All eleven log-

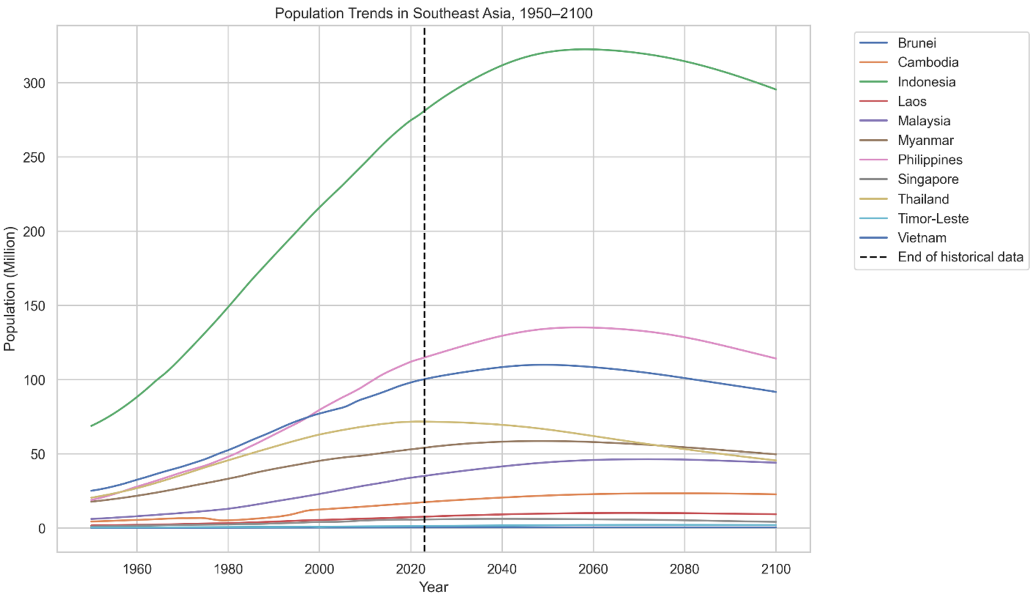

Figure 3. Population trends in Southeast Asia, 1950–2100.

Table 2. Residual diagnostic tests for final ARIMA models.

Country	ARIMA Order	ADF Residual	Ljung-Box	Jarque-Bera	ARCH-LM
Brunei	(1,1,2)	✓	✓	✗	✓
Cambodia	(1,1,1)	✓	✓	✗	✗
Indonesia	(1,1,1)	✓	✓	✗	✓
Laos	(2,1,1)	✓	✓	✗	✓
Malaysia	(1,1,1)	✓	✓	✗	✗
Myanmar	(1,1,1)	✓	✓	✗	✗
Philippines	(1,1,1)	✓	✓	✗	✗
Singapore	(1,1,2)	✓	✓	✗	✓
Thailand	(2,1,0)	✓	✓	✗	✓
Timor-Leste	(1,1,2)	✓	✓	✗	✓
Vietnam	(1,1,1)	✓	✓	✗	✓

transformed training series were found to be non-stationary in levels, as indicated by the ADF test, with p-values greater than 0.05 for all countries (Table 1). First-order differencing successfully achieved stationarity in eight countries, while three countries required second-order differencing. Visual inspection of the ACF and PACF plots further supported these findings, where the raw and log-transformed series exhibited persistent slow decay patterns. In contrast, the differenced series showed rapid decay toward zero within approximately two to three lags, indicating adequate stationarity after differencing.

Residual diagnostic results for the final ARIMA models are presented in Table 2. Overall, all eleven models satisfied the stationarity and white-noise assumptions based on the ADF and Ljung-Box tests, indicating that the fitted models adequately captured the autocorrelation structure of each series. However, residual normality was not achieved in any series according to the Jarque-Bera test, which is considered acceptable given the relatively limited training sample size.

Table 2 shows that homoscedasticity was satisfied in seven out of eleven models according to the ARCH-LM test. The remaining four countries, namely Cambodia, Malaysia, Myanmar, and the Philippines, exhibited residual variance clustering that may reflect demographic instability, policy-related disruptions, and rapid structural transitions not fully accommodated by logarithmic transformation. Despite these limitations, the overall diagnostic results indicate that the fitted ARIMA models were statistically adequate for short-term validation and benchmark comparison. However, their long-term outputs should still be interpreted as statistical extrapolations rather than complete demographic projections.

3.4. Stable Log-ARIMA Validation Performance

The Stable Log-ARIMA model was validated using historical population data from 1950 to 2010 for training and from 2011 to 2023 for validation. This validation was conducted to evaluate the model's short-term

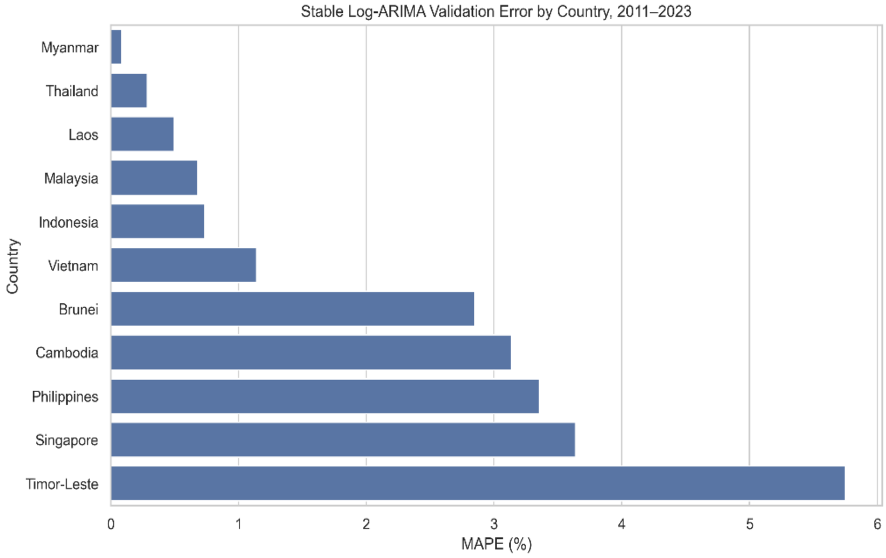

Figure 4. ARIMA validation error by country, 2011–2023.

forecasting performance before applying it to the long-term forecasting horizon.

Validation errors across Southeast Asian countries are summarized in Figure 4 using MAPE. The average validation MAPE was 2.01%, indicating that the model achieved relatively good short-term forecasting performance. The lowest validation error was observed in Myanmar, with a MAPE of 0.09%, while the highest error was observed in Timor-Leste, with a MAPE of 5.75%.

The country-level validation results in Table 3 provide the selected ARIMA orders, AIC values, and forecasting accuracy metrics for the 2011–2023 validation period. The selected ARIMA orders were determined by selecting the model with the lowest Akaike Information Criterion (AIC) for each country, indicating the most parsimonious model within the candidate set. Overall, the model produced strong short-term forecasting performance, with an average validation MAPE of 2.01%.

The lowest validation error was observed in Myanmar, with a MAPE of 0.09%, followed by Thailand, Laos, Malaysia, and Indonesia, all of which recorded MAPE values below 1%. These results indicate that the model reproduced short-term population dynamics with low relative error across several countries. Although Indonesia had relatively larger MAE and RMSE in absolute terms, this is expected because it has the largest population in the region. Therefore, MAPE provides a more comparable measure of relative forecasting accuracy across countries with different population scales.

Higher validation errors were observed in Timor-Leste, Singapore, the Philippines, Cambodia, and Brunei. Timor-Leste recorded the highest MAPE at 5.75%, suggesting that smaller or more irregular population series may be more difficult to forecast using a univariate ARIMA framework. Nevertheless, all country-level MAPE values remained below 6%, supporting the usefulness of Stable Log-ARIMA as a short-term statistical baseline for population forecasting.

However, low validation error over the 2011–2023 period does not necessarily imply reliable long-term forecasting performance. Long-term demographic trajectories are influenced not only by historical population values, but also by fertility decline, mortality improvement, migration, age structure, and broader demographic transition processes. Since these structural factors are not explicitly included in a univariate Stable Log-ARIMA model, the long-term forecasts should be interpreted as statistical extrapolations rather than full demographic projections.

3.5. Regional Comparison Between Stable Log-ARIMA Forecasts and UN WPP projections

After validation, the Stable Log-ARIMA model was used to forecast the population from 2024 to 2100. The forecast was then compared with the UN WPP Projections to evaluate the degree of alignment between a data-driven time-series model and a demographic projection benchmark.

The regional comparison in Figure 5 reveals that the Stable Log-ARIMA forecast and the UN WPP Projections are relatively close in the short-term horizon. In 2030, Stable Log-ARIMA estimates the regional population at approximately 724.96 million, while the UN WPP

Table 3. Stable Log-ARIMA validation performance by country.

Country	**ARIMA Order**	**AIC**	**MAE (million)**	**RMSE (million)**	**MAPE (%)**
Myanmar	(1,1,1)	-728.56	0.04	0.05	0.09
Thailand	(2,1,0)	-795.22	0.21	0.31	0.29
Laos	(2,1,1)	-743.73	0.04	0.05	0.49
Malaysia	(1,1,1)	-690.33	0.23	0.33	0.68
Indonesia	(1,1,1)	-697.51	2.03	3.01	0.74
Vietnam	(1,1,1)	-617.92	1.10	1.23	1.14
Brunei	(1,1,2)	-548.48	0.01	0.01	2.85
Cambodia	(1,1,1)	-366.36	0.53	0.70	3.14
Philippines	(1,1,1)	-621.87	3.74	4.81	3.35
Singapore	(1,1,2)	-504.92	0.21	2.86	3.64
Timor-Leste	(1,1,2)	-275.20	0.07	0.08	5.75
Average	-	-	-	-	**2.01**

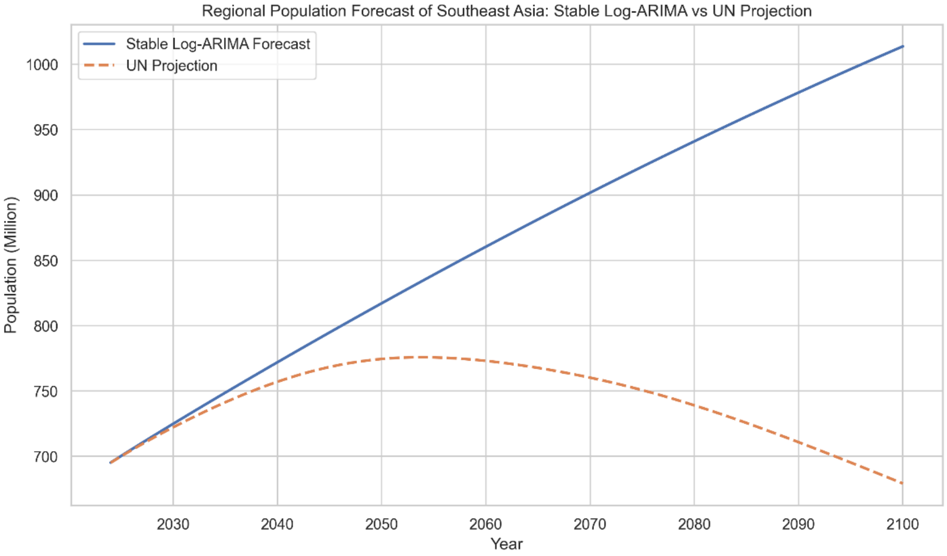

Figure 5. Regional population forecast of Southeast Asia: Stable Log-ARIMA versus UN WPP projections.

Table 4. Regional comparison between Stable Log-ARIMA forecasts and UN WPP projections.

Year	**Stable Log-ARIMA Forecast**	**UN WPP Projections**	**Difference (%)**
2030	724.96 million	722.29 million	0.37
2050	817.16 million	774.57 million	5.5
2075	933.41 million	727.42 million	28.32
2100	1,013.72 million	679.27 million	49.24

Projections estimates 722.29 million. This results in a small difference of approximately 0.37%.

However, the difference between the two trajectories becomes larger over time. In 2050, Stable Log-ARIMA estimates the regional population at 817.16 million, compared with 774.57 million from the UN WPP Projections. By 2100, Stable Log-ARIMA estimated the regional population at 1,013.72 million, while the UN WPP Projections estimated only 679.27 million.

The numerical comparison in Table 4 confirms that the difference between Stable Log-ARIMA and UN WPP Projections increases as the forecasting horizon becomes longer. The model remains close to the UN WPP Projections in the short term, but it increasingly overestimates population in the long term. The benchmark comparison between Stable Log-ARIMA and UN WPP Projections produced a regional alignment MAPE of 16.73% over the 2024–2100 period.

This result suggests that historical trend-based forecasting models tend to continue past population growth patterns. In contrast, the UN WPP Projections incorporates demographic assumptions such as fertility decline, mortality change, migration, and population aging. Therefore, Stable Log-ARIMA is suitable as a data-

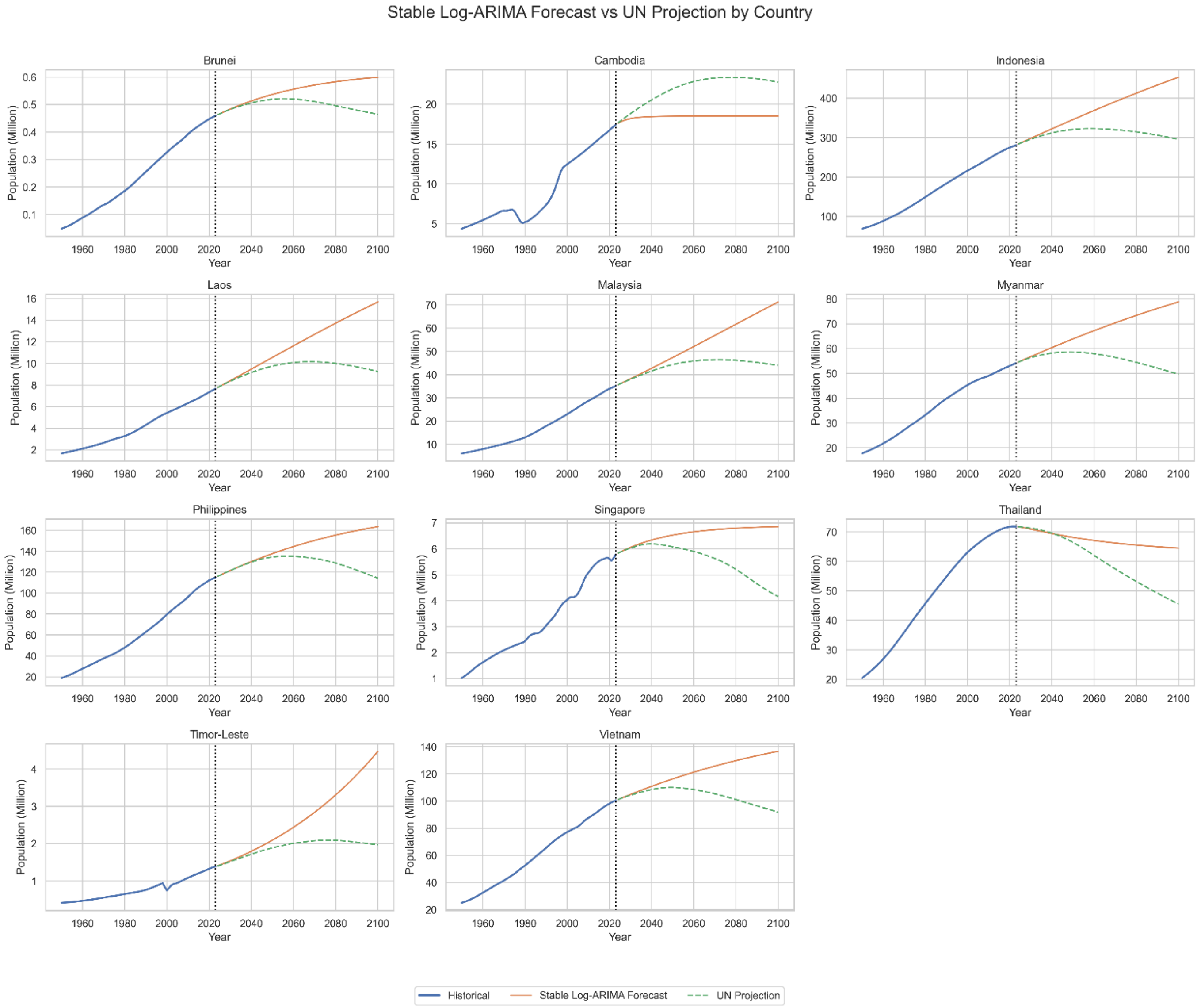

Figure 6. Country-level comparison between Stable Log-ARIMA forecast and UN WPP projections.

driven baseline model, but it should not be interpreted as a complete demographic projection model.

3.6. Country-level Comparison Between Stable Log-ARIMA and UN WPP Projections

The country-level comparison provides a more detailed view of how the Stable Log-ARIMA forecast differs from the UN WPP Projections across Southeast Asian countries. While the regional comparison shows overall divergence, country-level analysis helps identify which countries contribute most to the difference between the model-based forecast and the UN WPP Projections.

Country-level differences between historical population, Stable Log-ARIMA forecasts, and UN WPP Projections are displayed in Figure 6. The figure demonstrates that Stable Log-ARIMA generally produces smooth extrapolative trajectories based on historical population patterns. In the short term, the model trajectories remain relatively close to the UN WPP Projections for several countries. However, divergence becomes more visible over longer horizons, especially after the middle of the century.

Stable Log-ARIMA tends to project continued population growth in several countries, including Indonesia, Laos, Malaysia, Myanmar, the Philippines, Timor-Leste, and Vietnam. In contrast, the UN WPP Projections indicate that several of these countries are expected to reach a population peak and subsequently decline before 2100. The difference is particularly visible in Indonesia, Malaysia, Myanmar, Singapore, Thailand, and Vietnam. These differences highlight the limitation of univariate forecasting models, which rely only on historical population values and do not explicitly capture demographic transition mechanisms such as fertility decline, mortality improvement, migration, and age-structure change.

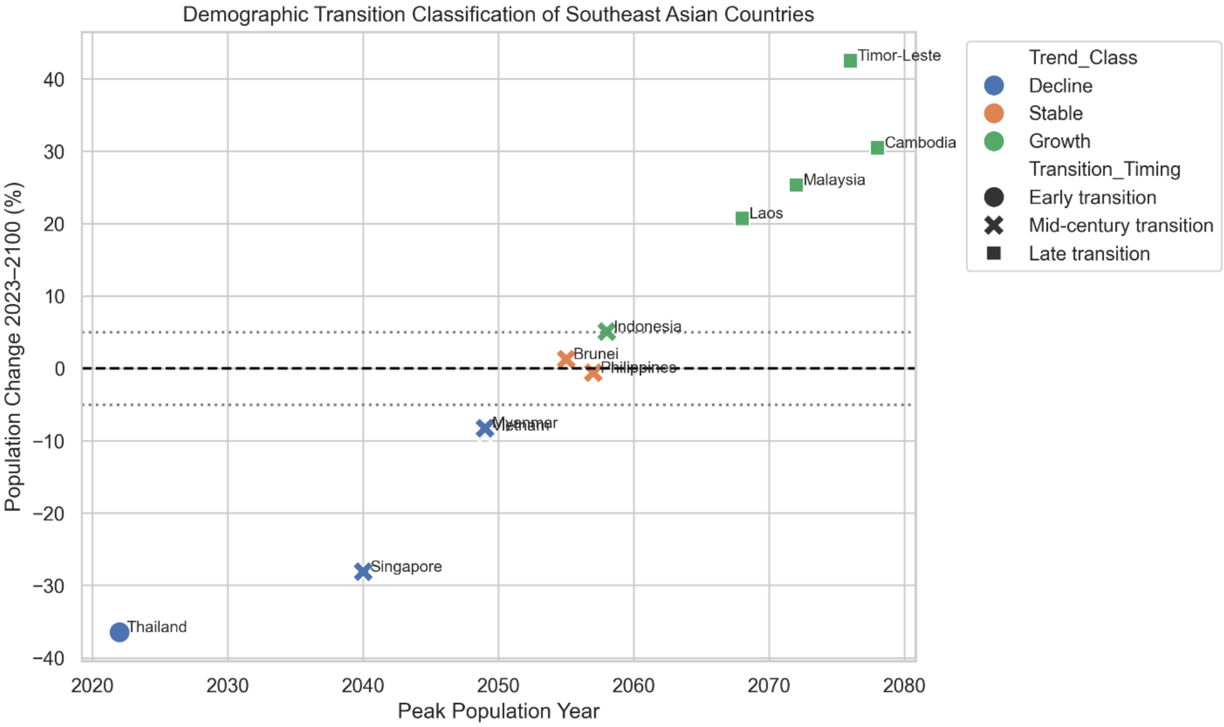

Figure 7. Demographic transition classification based on UN WPP projections.

Table 5. Demographic transition classification based on UN WPP projections.

Country	**Population 2023**	**Population 2050**	**Population 2100**	**Change 2023 2100 (%)**	**Peak Year**	**Trend Class**	**Transition Timing**
Thailand	71.7	66.38	45.56	-36.47	2022	Decline	Early transition
Singapore	5.79	6.08	4.16	-28.09	2040	Decline	Mid-century transition
Vietnam	100.35	110.01	91.71	-8.62	2049	Decline	Mid-century transition
Myanmar	54.13	58.62	49.66	-8.27	2049	Decline	Mid-century transition
Brunei	0.46	0.52	0.46	1.27	2055	Stable	Mid-century transition
Philippines	114.89	134.37	114.22	-0.58	2057	Stable	Mid-century transition
Indonesia	281.19	320.71	295.51	5.09	2058	Growth	Mid-century transition
Laos	7.67	9.76	9.26	20.76	2068	Growth	Late transition
Malaysia	35.13	44.29	44.03	25.34	2072	Growth	Late transition
Timor-Leste	1.38	1.89	1.97	42.52	2076	Growth	Late transition
Cambodia	17.42	21.93	22.74	30.5	2078	Growth	Late transition

3.7. Demographic transition classification

To provide a more interpretable demographic analysis, Southeast Asian countries were classified based on UN-projected population change from 2023 to 2100 and on their projected peak-population year. Countries with population increases of more than 5% were classified as Growth, countries with changes between -5% and +5% were classified as Stable, and countries with declines greater than 5% were classified as Decline.

The demographic transition map in Figure 7 shows that Southeast Asian countries are distributed across different stages of the transition. Thailand is classified as an early-transition country because its population peaked around 2022 and is projected to decline substantially by 2100. Singapore, Vietnam, and Myanmar are classified as declining countries with mid-century transition patterns.

Indonesia, Laos, Malaysia, Timor-Leste, and Cambodia are classified as growth countries. Indonesia is close to the growth threshold, while Timor-Leste and Cambodia show stronger long-term growth. Brunei and the Philippines are classified as stable because their projected population changes from 2023 to 2100 remain close to zero.

The classification results in Table 5 confirm that Southeast Asia does not follow a single demographic pathway. Thailand has already entered an early phase of decline, while Singapore, Vietnam, and Myanmar are projected to decline after reaching their mid-century population peaks. Brunei and the Philippines remain relatively stable by 2100. Meanwhile, Laos, Malaysia, Timor-Leste, and Cambodia are projected to continue growing for longer, indicating later demographic transitions.

This classification is one of the study's main contributions. Instead of presenting only numerical forecasts, the study provides a framework for interpreting long-term population change as growth, stability, or decline. This makes the results more useful for demographic interpretation and policy-oriented discussion.

3.8. Synthesis of Findings, Implications, and Limitations

The findings answer the main research question by showing that Stable Log-ARIMA is effective as a short-term statistical forecasting baseline but less reliable as a long-term demographic projection tool. During the 2011–2023 validation period, the model achieved an average MAPE of 2.01%, with several countries producing errors below 1%, indicating strong short-term accuracy. However, this accuracy does not extend uniformly to the long horizon. The model remains close to the UN WPP Projections in 2030, estimating 724.96 million compared with the UN estimate of 722.29 million, but diverges substantially by 2100, projecting 1,013.72 million compared with 679.27 million from the UN WPP Projections. This apparent conflict between short-term accuracy and long-term divergence is expected because Stable Log-ARIMA extrapolates historical population patterns, whereas the UN WPP Projections incorporates demographic mechanisms such as fertility decline, mortality improvement, migration, population aging, and age-structure change.

The demographic transition classification further shows that Southeast Asia does not follow a single demographic pathway. Thailand is already in an early decline phase; Singapore, Vietnam, and Myanmar show mid-century decline patterns; Brunei and the Philippines remain relatively stable; while Laos, Malaysia, Timor-Leste, and Cambodia are projected to continue growing for a longer period. This heterogeneity is important because regional forecasts can mask country-specific demographic transitions. The novelty of this study lies in combining Stable Log-ARIMA validation, long-term UN benchmark comparison, and demographic transition classification across all 11 Southeast Asian countries within a single integrated framework.

These findings are consistent with previous forecasting studies showing that ARIMA-based models can provide reliable short-term forecasts. Still, they may become less suitable for long-horizon projection when structural drivers are not explicitly modeled [8, 35]. In the context of population forecasting, this study extends prior evidence by showing that short-term validation accuracy alone is insufficient to justify long-term demographic reliability. Unlike cohort-component and probabilistic population projection approaches, which incorporate assumptions about fertility, mortality, migration, and age structure, Stable Log-ARIMA relies only on historical population trajectories. Therefore, the observed divergence from UN WPP projections reinforces the need to interpret univariate ARIMA forecasts as statistical baselines rather than substitutes for structurally informed demographic projections.

This study has several limitations. First, the model uses only the total population as a univariate time series variable. Second, the comparison with UN WPP Projections serves as a benchmark alignment assessment rather than an actual accuracy test, as future observed values are unavailable. Third, Stable Log-ARIMA does not explicitly incorporate fertility, mortality, migration, or age-structure indicators. Nevertheless, the findings clarify where a simple statistical model remains useful and where structurally informed demographic projection models become necessary. Future research should extend this framework by incorporating demographic covariates and comparing Stable Log-ARIMA with multivariate, Prophet, LSTM, and hybrid forecasting models.

4. Conclusions

This study addressed the research question by evaluating whether Stable Log-ARIMA can provide reliable population forecasts for Southeast Asia and how its long-term projections compare with UN World Population Prospects estimates. The results show that Stable Log-ARIMA performs well as a short-term statistical baseline, achieving an average validation MAPE of 2.01% during the 2011–2023 validation period. However, an important and partly unexpected finding is that strong short-term accuracy does not translate into reliable long-term demographic projection. Although the model remains close to the UN WPP Projections in the short term, it increasingly overestimates the regional population toward 2100, reflecting a discrepancy between univariate time-series extrapolation and structurally informed demographic projections. This divergence arises because Stable Log-ARIMA extrapolates historical population patterns, whereas UN WPP projections account for fertility decline, mortality improvement, migration, population aging, and age-structure dynamics. The demographic transition classification further confirms that Southeast Asia does not follow a single pathway, with early decline in Thailand, mid-century decline in Singapore, Vietnam, and Myanmar, relative stability in

Brunei and the Philippines, and continued growth in Laos, Malaysia, Timor-Leste, and Cambodia. The novelty of this study lies in integrating Stable Log-ARIMA validation, long-term UN benchmark comparison, and demographic transition classification across all eleven Southeast Asian countries. These findings imply that Stable Log-ARIMA is useful for transparent short-term forecasting and as a benchmark model. Still, it should not replace demographic projection models that incorporate structural population components. Future research should extend this framework by including fertility, mortality, migration, and age-structure indicators, and by comparing Stable Log-ARIMA with multivariate, Prophet, LSTM, and hybrid forecasting approaches.

Author Contributions: Conceptualization, G.M.I., F.A.H., and M.Z.A.; methodology, F.A.H.; software, G.M.I. and F.A.H.; validation, G.M.I. and F.A.H.; formal analysis, G.M.I.; investigation, G.M.I., F.A.H., and M.Z.A.; data curation, G.M.I.; writing original draft preparation, M.Z.A. for the Introduction, F.A.H. for the Materials and Methods, and G.M.I. for the Results and Discussion; writing review and editing, G.M.I., F.A.H., and M.Z.A.; visualization, G.M.I. and F.A.H; supervision, F.A.H.; project administration, G.M.I. All authors have read and agreed to the published version of the manuscript.

Funding: This study does not receive external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are publicly available from the Our World in Data population dataset, which compiles historical population estimates and United Nations World Population Prospects projections.

Acknowledgments: The authors acknowledge the publicly available population data provided by Our World in Data and the United Nations World Population Prospects.

Conflicts of Interest: All the authors declare no conflicts of interest.

References

1. Lee, R. (2025). Forecasting Population in an Uncertain World: Approaches, New Uses, and Troubling Limitations, *Population and Development Review*, Vol. 51, No. 1, 491–518. doi:10.1111/padr.12674.
2. Bhattacharjee, N. V, Schumacher, A. E., Aali, A., Abate, Y. H., Abbasgholizadeh, R., Abbasian, M., Abbasi-Kangevari, M., Abbastabar, H., Abd ElHafeez, S., Abd-Elsalam, S., Abdollahi, M., Abdollahifar, M.-A., Abdoun, M., Abdullahi, A., Abebe, M., Abebe, S. S., Abiodun, O., Abolhassani, H., Abolmaali, M., et al. (2024). Global Fertility in 204 Countries and Territories, 1950–2021, with Forecasts to 2100: A Comprehensive Demographic Analysis for the Global Burden of Disease Study 2021, *The Lancet*, Vol. 403, No. 10440, 2057–2099. doi:10.1016/S0140-6736(24)00550-6.
3. Jean Yeung, W.-J., and Abalos, J. B. (2025). Social Determinants of Low Fertility in Asia: A Comparative Review of Trends in East, Southeast and South Asia, *The Lancet Regional Health - Western Pacific*, Vol. 65, 101724. doi:10.1016/j.lanwpc.2025.101724.
4. Sakamoto, R. (2025). Aging in Southeast Asia and Japan: Challenges and Opportunities, *Geriatrics & Gerontology International*, Vol. 25, No. 7, 837–854. doi:10.1111/ggi.70062.
5. Wiśniowski, A., and Raymer, J. (2025). Multiregional Population Forecasting: A Unifying Probabilistic Approach for Modelling the Components of Change, *European Journal of Population*, Vol. 41, No. 1, 11. doi:10.1007/s10680-025-09729-7.
6. Mohan, A., Jaihind Jothikaran, T. A., Gudi, N., A K, A., and Ashok, L. (2025). Growing Old before Growing Rich: A Scoping Review on Ageing-Related Policy Interventions from the WHO Southeast Asian Region, *BMJ Open*, Vol. 15, No. 1, e079621. doi:10.1136/bmjopen-2023-079621.
7. United Nations Population Division, D. of E. and S. A. (2022). *World Population Prospects 2022: Methodology of the United Nations Population Estimates and Projections*New York, United Nations.
8. Raftery, A. E., and Ševčíková, H. (2023). Probabilistic Population Forecasting: Short to Very Long-Term, *International Journal of Forecasting*, Vol. 39, No. 1, 73–97. doi:10.1016/j.ijforecast.2021.09.001.
9. Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., and Matsopoulos, G. K. (2023). A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks, *Future Internet*, Vol. 15, No. 8, 255. doi:10.3390/fi15080255.
10. Taslim, D. G., and Murwantara, I. M. (2024). Comparative Analysis of ARIMA and LSTM for Predicting Fluctuating Time Series Data, *Bulletin of Electrical Engineering and Informatics*, Vol. 13, No. 3, 1943–1951. doi:10.11591/eei.v13i3.6034.
11. Ab Rashid, M. A., Ahmad Zaki, R., Wan Mahiyuddin, W. R., and Yahya, A. (2023). Forecasting New Tuberculosis Cases in Malaysia: A Time-Series Study Using the Autoregressive Integrated Moving Average (ARIMA) Model, *Cureus*, Vol. 15, No. 9, e44676. doi:10.7759/cureus.44676.
12. Schaffer, A. L., Dobbins, T. A., and Pearson, S.-A. (2021). Interrupted Time Series Analysis Using Autoregressive Integrated Moving Average (ARIMA) Models: A Guide for Evaluating Large-Scale Health Interventions, *BMC Medical Research Methodology*, Vol. 21, No. 1, 58. doi:10.1186/s12874-021-01235-8.
13. Wilson, T., Grossman, I., Alexander, M., Rees, P., and Temple, J. (2022). Methods for Small Area Population Forecasts: State-of-the-Art and Research Needs, *Population Research and Policy Review*, Vol. 41, No. 3, 865–898. doi:10.1007/s11113-021-09671-6.
14. Lütkepohl, H., and Xu, F. (2012). The Role of the Log Transformation in Forecasting Economic Variables, *Empirical Economics*, Vol. 42, No. 3, 619–638. doi:10.1007/s00181-010-0440-1.
15. Ma, J., Chen, Q., Chen, X., Fan, J., and Li, X. (2025). An Inevitably Ageing World: Analysis on the Evolutionary Pattern of Age Structure in 200 Countries, *Royal Society Open Science*, Vol. 12, No. 4. doi:10.1098/rsos.241988.
16. Ritchie, H., Rosado, P., and Roser, M. (2023). Population Growth, *Our World in Data*.
17. Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., et al. (2020). Array Programming with NumPy, *Nature*, Vol. 585, No. 7825, 357–362. doi:10.1038/s41586-020-2649-2.
18. Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2018). The M4 Competition: Results, Findings, Conclusion and Way Forward, *International Journal of Forecasting*, Vol. 34, No. 4, 802–808. doi:10.1016/j.ijforecast.2018.06.001.
19. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.), Wiley, Hoboken, NJ.

20. Hyndman, R. J., and Khandakar, Y. (2008). Automatic Time Series Forecasting: The Forecast Package for R, *Journal of Statistical Software*, Vol. 27, No. 3, 1–22. doi:10.18637/jss.v027.i03.

21. Granger, C. W. J., and Newbold, P. (1986). *Forecasting Economic Time Series* (2nd ed.), Academic Press, Orlando, FL.

22. Brockwell, P. J., and Davis, R. A. (2016). *Introduction to Time Series and Forecasting* (3rd ed.), Springer International Publishing, Cham. doi:10.1007/978-3-319-29854-2.

23. Seabold, S., and Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python, *Proceedings of the 9th Python in Science Conference*, 92–96. doi:10.25080/Majora-92bf1922-011.

24. Hyndman, R. J., and Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.), OTexts, Melbourne.

25. Dickey, D. A., and Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association*, Vol. 74, No. 366a, 427–431. doi:10.1080/01621459.1979.10482531.

26. Akaike, H. (1974). A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control*, Vol. 19, No. 6, 716–723. doi:10.1109/TAC.1974.1100705.

27. Melina, M., Sukono, S., Napitupulu, H., Mohamed, N., Chrisnanto, Y. H., Hadiana, A. I., Kusumaningtyas, V. A., and Nabilla, U. (2024). Comparative Analysis of Time Series Forecasting Models Using Arima and Neural Network Autoregression Methods, *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, Vol. 18, No. 4, 2563–2576. doi:10.30598/barekengvol18iss4pp2563-2576.

28. LJUNG, G. M., and BOX, G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models, *Biometrika*, Vol. 65, No. 2, 297–303. doi:10.1093/biomet/65.2.297.

29. Jarque, C. M., and Bera, A. K. (1980). Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals, *Economics Letters*, Vol. 6, No. 3, 255–259. doi:10.1016/0165-1765(80)90024-5.

30. Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, Vol. 50, No. 4, 987. doi:10.2307/1912773.

31. Hyndman, R. J., and Koehler, A. B. (2006). Another Look at Measures of Forecast Accuracy, *International Journal of Forecasting*, Vol. 22, No. 4, 679–688. doi:10.1016/j.ijforecast.2006.03.001.

32. Makridakis, S., Wheelwright, S. C., and Hyndman, R. J. (1998). *Forecasting: Methods and Applications* (3rd ed.), Wiley, New York.

33. Reher, D. S. (2004). The Demographic Transition Revisited as a Global Process, *Population, Space and Place*, Vol. 10, No. 1, 19–41. doi:10.1002/psp.313.

34. Gerland, P., Raftery, A. E., Ševčíková, H., Li, N., Gu, D., Spoorenberg, T., Alkema, L., Fosdick, B. K., Chunn, J., Lalic, N., Bay, G., Buettner, T., Heilig, G. K., and Wilmoth, J. (2014). World Population Stabilization Unlikely This Century, *Science*, Vol. 346, No. 6206, 234–237. doi:10.1126/science.1257469.

35. Swanson, D. A., and Tayman, J. (2025). On Using ARIMA Model Confidence Intervals Applied to Population Projections Based on the Components of Change: A Case Study for the World Population, *Statistics in Transition New Series*, Vol. 26, No. 4, 1–14. doi:10.59139/stattrans-2025-036.