APPLYING THE ARIMAX MODEL TO FORECAST POPULATION CHANGES: A STUDY IN DANANG CITY AFTER THE MERGER

VẬN DỤNG MÔ HÌNH ARIMAX DỰ BÁO BIẾN ĐỘNG DÂN SỐ: NGHIÊN CỨU TẠI THÀNH PHỐ ĐÀ NẴNG SAU HỢP NHẤT

Nguyen Ba The, Nguyen Van Long, Doan Ngoc Phi Anh*

The University of Danang - University of Economics, Vietnam

*Corresponding author: phianhdn@due.edu.vn

(Received: May 08, 2026; Revised: May 25, 2026; Accepted: June 19, 2026)

DOI: 10.31130/ud-jst.2026.24(6B).321E

Abstract - Population is a fundamental factor that plays a central role in socio-economic development at both national and local levels. Changes in population size and structure not only affect the labor force but also have significant implications for social welfare and economic growth, particularly in the context of the administrative integration between Da Nang and Quang Nam. This study employs descriptive statistical methods and the ARIMAX model to analyze population structure characteristics and forecast population changes in Da Nang City using data from 1999 to 2024. In addition, the study examines the effects of socio-economic factors on population dynamics and projects the population size for the 2025–2030 period. The results indicate that the ARIMAX(0,0,1) model incorporating sex ratio as an exogenous variable is statistically significant ($p < 0.05$) and achieves a forecast accuracy with a MAPE of 9.55%. The population of Da Nang is projected to reach approximately 3.27 million by 2030.

Key words - ARIMA Model; population structure; economy factors; volatility; forecast.

Tóm tắt – Dân số là yếu tố cốt lõi, giữ vai trò trung tâm trong sự phát triển kinh tế - xã hội ở cấp quốc gia và địa phương. Sự thay đổi về quy mô và cơ cấu dân số không chỉ ảnh hưởng đến nguồn lao động mà còn tác động đến an sinh xã hội và tăng trưởng kinh tế, đặc biệt trong tình hình hợp nhất không gian hành chính giữa Đà Nẵng và Quảng Nam. Nghiên cứu sử dụng phương pháp thống kê mô tả và mô hình ARIMAX để phân tích và dự báo đặc điểm cơ cấu dân số sử dụng dữ liệu từ năm 1999 đến 2024. Thêm vào đó, nghiên cứu lượng hóa mức độ cũng như chiều hướng tác động của các yếu tố kinh tế - xã hội đến biến động dân số, đồng thời dự báo quy mô dân số Thành phố Đà Nẵng trong giai đoạn 2025-2030. Kết quả cho thấy mô hình ARIMAX(0,0,1) với biến ngoại sinh là tỷ số giới tính có ý nghĩa thống kê ($p < 0,05$), đạt sai số dự báo MAPE ≈ 9,55%; dân số Đà Nẵng được dự báo đạt khoảng 3,27 triệu người vào năm 2030.

Từ khóa – Mô hình ARIMAX; cơ cấu dân số; yếu tố kinh tế; biến động; dự báo.

1. Introduction

Vietnam is currently entering a period of population aging, characterized by declining fertility rates and increasingly complex migration trends. In this context, monitoring and analyzing population change trends at the local level has become an urgent requirement to support planning and sustainable development policymaking. Vietnam's population reached approximately 101.1 million people in 2024; the total fertility rate declined to 1.91 children per woman, below the replacement level, while the proportion of people aged 65 and above reached 9.3%, and the aging index increased rapidly to 60.2%. In addition, the urban population grew at an average annual rate of 3.06%, significantly higher than in previous periods, reflecting strong urbanization and shifts in population structure [1]. These figures indicate that Vietnam's population structure is changing rapidly, creating an urgent need for local authorities to assess, forecast, and formulate appropriate population policies.

The city of Da Nang is an important socio-economic center of Central Vietnam, with a high rate of urbanization and economic development. The city hosts many key economic and industrial zones, such as Da Nang Hi-Tech Park, Hoa Khanh Industrial Park, and Lien Chieu Industrial Park. In addition, Da Nang is developing the Da Nang International Financial Center and the Da Nang Free Trade Zone to enhance investment attraction, promote trade and financial activities, and foster international economic integration. The development of these economic zones not only contributes to economic growth, job creation, and higher incomes for residents, but also significantly affects the size, structure, and spatial distribution of the population within the city. This process has led to noticeable changes in occupational structure, labor qualifications, migration flows from neighboring localities, and population distribution across wards and communes. However, there remains a lack of quantitative studies in Vietnam, including Da Nang City, that apply time-series models with exogenous variables to forecast population changes at the local level.

Therefore, measuring the impact of socio-economic factors on the city's population dynamics in the context of administrative boundary mergers, while also forecasting population changes in newly merged localities, is of significant importance.

2. Literature review

2.1. Conceptual framework

The population structure by gender reflects the proportional relationship between the number of males and females within a population at a given point in time. Differences between the two genders are a fundamental characteristic of population structure and have profound impacts on demographic processes as well as socio-economic development. To describe the population structure by gender, the most commonly used indicator is the sex ratio, which represents the number of males per 100 females in the total population. This indicator reflects the degree of balance or imbalance between the two genders within a community [2].

The urban–rural population structure refers to the ratio of the urban population to the rural population within the total population at a given point in time [2].

The overall population growth rate (often simply referred to as the population growth rate) reflects the change in population size of a given area over a specific period, including changes resulting from both natural increase and migration. This indicator is expressed either as a percentage (%) or per thousand (‰), depending on the purpose of analysis and comparison [2].

2.2. Empirical studies

Quantitative models, particularly time-series models, have demonstrated high effectiveness in analyzing and forecasting population trends. Among them, the ARIMA (Autoregressive Integrated Moving Average) model, based on the Box–Jenkins methodology, is a prominent tool due to its ability to capture the temporal dependency structure of data [3].

Empirical evidence has shown the effectiveness and broad applicability of the ARIMA model. A study by González-Leonardo et al. [4] in the Greater Mexico Metropolitan Area used data from the 2000–2010 period to conduct retrospective forecasts for 2011–2020, achieving a median error of only about 1.5% for total population over a ten-year period. This result confirms the high accuracy of the ARIMA model even at small geographical scales, while also demonstrating its potential for application at the provincial/city level or lower.

The widespread use of the model has also been reinforced through numerous studies conducted in different contexts, such as the research of Nyoni [5] in Yemen and Russia, Zakria & Muhammad [6] in Pakistan, as well as Dai & Chen [7] and Gu [8] in China. Despite differences in socio-economic conditions, these studies consistently indicate that ARIMA is capable of effectively identifying and extrapolating population trends under relatively stable conditions, thereby providing a reliable scientific basis for forecasting and planning for policymakers. Building upon these theoretical arguments and prior empirical evidence, this study employs an ARIMAX model instead of the traditional ARIMA approach in order to simultaneously capture the dynamic properties of the time series and the effects of exogenous variables, namely the sex ratio and the urban–rural population ratio. The inclusion of these independent variables enables the model not only to generate forecasts based on historical population trends but also to account for the impact of structural demographic changes and urbanization processes on population growth in Da Nang City.

2.3. The relationship between population growth rate, sex ratio, and urban–rural population ratio

The sex ratio is examined within the frameworks of demographic transition theory, the economics of fertility behavior, and socio-cultural theories [9], [10], [11]. This variable reflects the gender structure of the population and influences marriage patterns [12]. When gender imbalance occurs, marriage opportunities for a portion of the population may decline [9], [13]. This situation leads to lower fertility rates and slows population growth [12], [13]. Therefore, the sex ratio indirectly affects population growth through the marriage market mechanism [11], [14].

The urban–rural ratio is examined within the frameworks of urbanization theory and demographic transition theory [10], [11]. This factor reflects the level of urbanization of the economy, thereby influencing fertility behavior and migration patterns of the population [13]. According to household economics theory, fertility decisions are affected by child-rearing costs and opportunities for labor market participation, particularly for women [9]. Urbanization is often associated with higher living costs, better educational attainment, and increased female labor force participation, which contribute to lower fertility rates and slower population growth [8], [11]. Therefore, the urban–rural ratio is expected to have a negative impact on the population growth rate [14].

2.4. Proposed model

The study employs the ARIMA (Auto Regressive Integrated Moving Average) model with exogenous variables, namely the ARIMAX (ARIMA with eXogenous variables) model, in which:

- The dependent variable is Tltds: Population growth rate of Da Nang City (%)

- The independent variables include:

+ TsoGt: Sex ratio of Da Nang City (%);

+ TsoTtNt: Urban–rural population ratio of Da Nang City (%).

Using data on these variables for Da Nang City after the merger from the two provinces, the study identifies several characteristics of the variables. Specifically:

Variable Tltds (population growth rate): Observation of Table 1 shows that the TLtds series exhibits strong variance fluctuations and experienced significant shocks during the 2015–2016 period. However, to accurately assess stationarity, the study conducts both ADF and PP unit root tests. The statistical results (shown in Table 1) indicate that the p-values of both tests are 0.0000 (smaller than the 0.01 significance level), leading to the rejection of the null hypothesis of a unit root and confirming that the Tltds series is actually stationary at level, I(0). Therefore, for inclusion in the ARIMAX model, the study only applies a logarithmic transformation to smooth the variance and minimize the impact of outliers, without the need to perform first-order differencing.

Table 1. *Components of the Proposed Model*

Observed Variable	ADF - Level	PP - Level	ADF - First Difference	PP - First Difference	Stationarity Conclusion
Tltds	-6.3214*** (0.0000)	-6.3214*** (0.0000)	-	-	Stationary at level I(0)

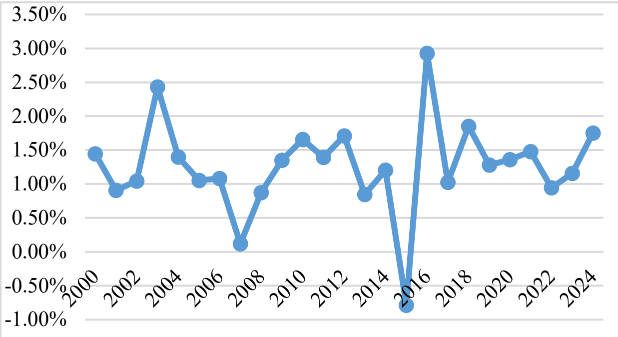

Figure 1. *Population Growth Rate*

The variable TsoGt (sex ratio) exhibits a notable long-term upward trend, particularly after 2017, as illustrated in Figure 2. The mean of the series appears to vary over time rather than remaining constant, suggesting potential instability in its underlying process. Moreover, the series demonstrates several pronounced structural breaks, including a sharp decline in 2010, a sudden increase in 2015, and another marked decrease in 2017. These fluctuations indicate the presence of heteroskedasticity. Specifically, the series shows relatively low volatility during the 2000–2009 period, followed by a substantial increase in volatility between 2010 and 2018, before gradually stabilizing thereafter. Although preliminary visual inspection suggests that the series may be non-stationary, both Augmented Dickey–Fuller (ADF) and Phillips–Perron (PP) unit root tests were conducted to provide a more rigorous assessment. The results reported in Table 2 indicate that the p-values for both tests are 0.0244, which is below the 5% significance level. Accordingly, the null hypothesis of a unit root is rejected, confirming that the TsoGt series is stationary at level, i.e., I(0). Based on these findings, the study applies only a natural logarithmic transformation (Ln) to the variable in the subsequent model estimation in order to stabilize variance and mitigate the influence of shocks or outliers. The application of an additional first-order differencing transformation is deemed unnecessary from a statistical standpoint and could result in the loss of important information embedded in the original data series.

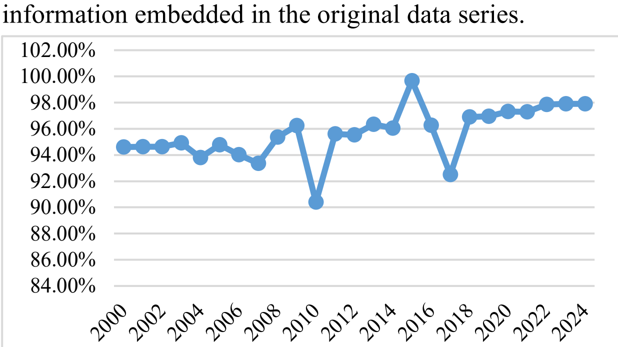

Figure 2. *Sex Ratio*

Table 2. *Results of the unit root tests*

Observed Variable	ADF - Level	PP - Level	ADF - First Difference	PP - First Difference	Stationarity Conclusion
TsoGt	-3.3346** (0.0244)	-3.3346** (0.0244)	-	-	Stationary at level I(0)

Variable TsoTtNt (urban–rural population ratio): Observation of Figure 3 shows that the TsoTtNt variable exhibits a clear and continuous upward trend over time, increasing from 94.69% in 2000 to 97.92% in 2024. This indicates that the mean value of the series changes over time and suggests a violation of the stationarity condition.

This visual observation is fully confirmed by the ADF and PP unit root tests. At the level form, the p-values of both tests are very high (0.9091 and 0.9724, respectively), indicating that the series is non-stationary. However, after taking the first difference, both tests produce p-values below 0.05, leading to the rejection of the null hypothesis (H0), thereby confirming that the series becomes stationary and is integrated of order one, denoted as I(1) as shown in Table 3.

Regarding variance, the data fluctuate relatively stably around the mean value, with no clear evidence of heteroskedasticity. Therefore, the necessary technical solution to transform the series into a stationary process is to apply first-order differencing. In addition, within the context of the final ARIMAX model, a natural logarithm (Ln) transformation is also applied to ensure consistency with the other variables in the model and to facilitate the interpretation of coefficients in terms of elasticity.

Table 3. *Results of the unit root tests*

Observed Variable	ADF - Level	PP - Level	ADF - First Difference	PP - First Difference	Stationarity Conclusion
TsoTtNt	-0.3130 (0.9091)	0.2866 (0.9724)	-5.2892*** (0.0003)	-7.4446*** (0.0000)	Stationary at first difference I(1)

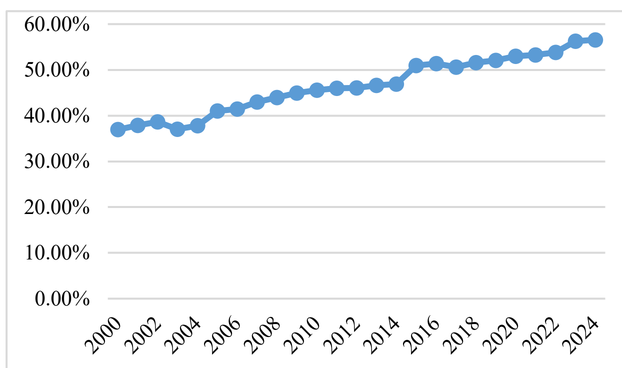

Figure 3. *Urban–Rural Population Ratio*

Based on the above diagnostic analyses, the proposed model can be specified as follows:

$$\mathrm{Ln}(\mathrm{Tltds})_t = \alpha + \sum_{i=1}^{p}\phi_i \mathrm{Ln}(\mathrm{Tltds})_{t-1} + \sum_{j=1}^{q}\theta_j\varepsilon_{t-j} + \sum_{m=0}^{r}\beta_k \mathrm{Ln}(\mathrm{TsoGt})_{t-m} + \sum_{n=0}^{s}\gamma_n \Delta \mathrm{Ln}(\mathrm{TsoTtNt})_{t-n} + \varepsilon_t$$

Table 4. *Components of the Proposed Model*

Component	Symbol	Explanation
Intercept	$\boldsymbol{\alpha}$	
Autoregressive (AR)	$\sum_{i=1}^{p}\phi_i Ln(Tltds)_{t-i}$	*The general form of the AR(p) component (order p), describing the influence of lagged values (or past values) of the dependent variable itself.*
Moving Average (MA)	$\sum_{j=1}^{q}\theta_j\varepsilon_{t-j}$	*General form of the MA(q) component (order q). Describes the influence of past error terms.*
Exogenous Variable 1	$\sum_{m=0}^{r}\beta_m Ln(TsoGt)_{t-m}$	*Describes the distributed lag effect of TsoGt from Lag 0 to Lag r.*
Exogenous Variable 2	$\sum_{n=0}^{s}\gamma_n \Delta Ln(TsoTtNt)_{t-n}$	*Describes the distributed lag effect of the first difference of the TsoTtNt variable from Lag 0 to Lag s.*
Error	$\boldsymbol{\varepsilon_t}$	*White noise (innovation) error term at time t.*

3. Research methods

3.1. Data

The study primarily employs secondary data collected from official statistical sources, thereby ensuring reliability, completeness, and consistency over time. The statistical data sources include the *Vietnam Statistical Yearbook*, the *Statistical Yearbooks of Quang Nam Province and Da Nang City* (1999–2024), reports from the *Population and Housing Census* (1999, 2009, and 2019), as well as the *Results of the 2024 Mid-term Population Survey* published by General Statistics Office of Vietnam.

3.2. Data Processing

To achieve the research objectives and address the research contents, the study employs both descriptive statistics and inferential statistics methods. Specifically:

- Period 1999–2024: Descriptive statistical methods, including charts, cross-tabulations, and trend lines, are utilized to illustrate the characteristics of population structure over time, by gender, and by area of residence.

- Long population data series for the period 1999–2024: Inferential statistical methods, particularly the ARIMAX model, are employed to quantify the magnitude and direction of the impacts of socio-economic factors (sex ratio and urban–rural population ratio) on population fluctuations. In addition, the model is used to forecast the population size of Da Nang for the period 2025–2030.

4. Research results

4.1. Population size

During the period 1999–2009, the average annual population growth rate of Da Nang was 1.17% per year, corresponding to an average increase of approximately 25.44 thousand people annually. Entering the 2009–2019 period, the population growth rate increased significantly, averaging 1.31% per year, equivalent to an additional increase of about 31.2 thousand people each year. However, during the 2019–2024 period, the population growth rate showed signs of stabilization while still maintaining an upward trend, with an average growth rate of 1.33% per year and an average annual increase of 36.28 thousand people, as presented in Table 5.

Table 5. *Average population of Da Nang after the administrative merger*

Indicator	1999	2009	2019	2024 (Estimated)
Total population (thousand persons)	2063.1	2317.5	2638.5	2819.9
Average annual population growth rate (%)	-	1.17	1.31	1.33
Average annual absolute increase (decrease) (thousand persons)	-	25.44	31.20	36.28

Source: Vietnam Statistical Yearbook and the author's calculations.

4.2. Population structure

4.2.1. Population structure by gender

The sex ratio is an indicator reflecting the level of balance between males and females in the total population. During the 1999–2009 period, the sex ratio in Da Nang fluctuated between 94.69 and 95.9 males per 100 females, indicating a surplus of females. In the 2009–2019 period, the ratio showed improvement, ranging from 95.9 to 96.99 males per 100 females; however, it still reflected a predominance of females. This trend continued to improve during the 2019–2024 period, with the sex ratio fluctuating between 96.69 and 97.91 males per 100 females. Compared with the national average, the sex ratio in Da Nang has increased steadily over the years; nevertheless, the city still exhibits a degree of gender imbalance, ranging from 94.69 to 97.91 males per 100 females, as presented in Table 6.

Table 6. *Population structure of Da Nang after the administrative merger*

	1999		2009		2019		2024 (Estimated)	
By gender	**1,000 persons**	**%**	**1,000 persons**	**%**	**1,000 persons**	**%**	**1,000 persons**	**%**
* Male	1003.4	48.6	1134.5	49.0	1298.9	49.2	1395.1	49.5
* Female	1059.7	51.4	1183.0	51.0	1339.6	50.8	1424.8	50.5
Sex Ratio (%)	**94.69**		**95.9**		**96.96**		**97.91**	
By area	**1,000 persons**	**%**	**1,000 persons**	**%**	**1,000 persons**	**%**	**1,000 persons**	**%**
* Urban	738.9	35.8	1041	44.9	1374.2	52.1	1595.3	56.6
* Rural	1324.2	64.2	1276.5	55.1	1264.3	47.9	1224.6	43.4

4.2.2. Population structure by urban and rural areas

The population structure of Da Nang during the 1999–2024 period demonstrates significant changes between urban and rural areas. Throughout the 1999–2019 period, the urban population remained smaller than the rural population. In 1999, the urban population accounted for only 35.82% of the total population. By 2019, the proportion of the urban population had increased to 44.92%. Since 2019, the urbanization process in Da Nang has accelerated considerably, with the share of the urban population rising to 52.08% in 2019 and reaching 56.57% by 2024.

These changes indicate that Da Nang is undergoing a demographic transition toward urbanization, characterized by a steadily increasing proportion of the urban population. This trend not only reflects the impacts of economic development, infrastructure investment, and urban spatial expansion, but is also closely associated with labor migration from rural to urban areas in search of employment opportunities and more stable incomes.

4.3. Model estimation and forecasting results

To estimate the model parameters, the study employs the ARIMA model with exogenous variables (ARIMAX) using the SPSS software package. After handling the outliers, the results of the ARIMAX(0,0,1) model indicate that the variable TsoTtNt has a significance value (p-value) of 0.716, which is substantially higher than the conventional threshold of 0.05. This finding suggests that, based on the collected sample data, the TsoTtNt variable does not have a statistically significant effect on the dependent variable Tltds (see Figure 4).

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Tltds-Model_1	Tltds	No Transformation	Constant		-3.648	1.337	-2.729	.013
			MA	Lag 1	-.145	.216	-.669	.511
	TsoGt	No Transformation	Numerator	Lag 0	25.869	11.746	2.202	.039
	TsoTtNt	No Transformation	Numerator	Lag 0	-.728	1.972	-.369	.716

Figure 4. *Statistical Results of the ARIMA Model Estimating the Population Growth Rate of Da Nang (2000–2024)*

The estimation results of the ARIMA(0,0,1) model after excluding the TsoTtNt variable are presented in Figure 5.

Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	Ljung-Box Q(18): Statistics	DF	Sig.	Number of Outliers
Tltds-Model_1	1	.234	11.756	17	.815	0

Figure 5. *Model Statistics*

The results indicate that the Stationary R-squared value reaches 0.234 (see Figure 6), reflecting the degree of improvement of the model compared with the baseline model (without structural components) in forecasting the time series after trend removal. Although this value is not particularly high, within the context of time-series modeling, Stationary R-squared is not considered the primary evaluation criterion. The adequacy of the model is confirmed through residual diagnostics and forecasting accuracy measures. Specifically:

(1) The model does not violate fundamental assumptions (the most important criterion)

The Ljung–Box Q(18) test yields a significance value of 0.815. This very high value (> 0.05) indicates that the residuals (errors) of the model are entirely random white noise. This finding suggests that the model has successfully extracted all systematic patterns in the time series, without omitting any important signals, and does not suffer from autocorrelation, which is considered a serious issue in time-series analysis.

(2) Low forecasting error (highly satisfactory MAPE value)

The findings show that the ARIMAX(0,0,1) model achieves a Mean Absolute Percentage Error (MAPE) of approximately 9.55%, indicating a relatively good level of forecasting accuracy. This result is consistent with previous studies. For example, González-Leonardo et al. [4], when applying the ARIMA model to the urban area of Mexico City, reported a median long-term forecasting error of approximately 1.5%. Similarly, Al-Zraigat et al. [15] demonstrated that ARIMA/ARIMAX models can achieve high accuracy in forecasting demographic indicators such as birth rates. Although the forecasting error in the present study is relatively higher, this outcome is reasonable due to the smaller dataset and the greater population volatility at the local level compared with national or regional scales.

Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.234	.	.234	.234	.234	.234	.234	.234	.234	.234	.234
R-squared	.234	.	.234	.234	.234	.234	.234	.234	.234	.234	.234
RMSE	.961	.	.961	.961	.961	.961	.961	.961	.961	.961	.961
MAPE	9.547	.	9.547	9.547	9.547	9.547	9.547	9.547	9.547	9.547	9.547
MaxAPE	36.477	.	36.477	36.477	36.477	36.477	36.477	36.477	36.477	36.477	36.477
MAE	.554	.	.554	.554	.554	.554	.554	.554	.554	.554	.554
MaxAE	3.291	.	3.291	3.291	3.291	3.291	3.291	3.291	3.291	3.291	3.291
Normalized BIC	.307	.	.307	.307	.307	.307	.307	.307	.307	.307	.307

Figure 6. *Model Fit*

(3) The independent variable TsoGt has a strong impact

The model identifies TsoGt as a significant explanatory variable affecting TlTds. With a significance value of 0.024 (< 0.05), this variable has a positive and statistically significant effect, with an estimated coefficient as high as 23.522 (as shown in Figure 7). This finding indicates that the study appropriately incorporated this variable into the analysis, as it substantially explains the variation in the dependent variable.

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Tltds-Model_1	Tltds	No Transformation	Constant		-3.190	.487	-6.553	.000
			MA	Lag 1	-.154	.211	-.728	.474
	TsoGt	No Transformation	Numerator	Lag 0	23.522	9.690	2.427	.024

Figure 7. *Variables in the ARIMA Model*

The result is consistent with previous studies, such as those of Becker [9] and Dyson [11], which suggest that gender structure influences marriage and fertility behavior through marriage market mechanisms. In addition, the study by Chipman and Morrison [12] also indicated that gender imbalance may affect birth rates at the local level. These findings demonstrate that incorporating exogenous variables into the ARIMAX model enhances its explanatory power compared with the traditional ARIMA model.

(4) The data are stable and free from extreme noise

The Number of Outliers index equals 0, indicating that the dataset does not contain abnormal fluctuations or excessively large measurement errors that could distort the results. Therefore, the obtained results are highly representative and accurately reflect the underlying patterns of the phenomenon under investigation.

The model can be expressed as follows:

$$\mathrm{Ln}(\mathrm{Tltds})_t = -3.190 + 23.522 * \mathrm{Ln}(\mathrm{TsoGt}_t) - 0.154\mathrm{e}_{t-1} + \mathrm{e}_t$$

The estimated model follows a Log–Log structure. In this model, the estimated coefficient β = 23.522 for the variable Ln(TsoGt) is interpreted as an elasticity coefficient. Specifically, when the sex ratio increases by 1%, the population growth rate increases by approximately 23.522%.

At first glance, the coefficient value of 23.522 may appear relatively large. However, from the mathematical perspective of demographic data, this magnitude is entirely reasonable. This is because the population growth rate (Tltds) itself has a very small baseline value (fluctuating around 1%–2%). Therefore, a relative increase of 23.522% from a 1% baseline actually corresponds to an increase of only about 0.2352 percentage points (for example, from 1.00% to 1.235%).

This actual impact magnitude (0.23522 percentage points) is fully consistent with the natural demographic fluctuations observed in Da Nang City, thereby explaining the seemingly large estimated coefficient. Using this model, the population of Da Nang is forecasted as presented in Table 7.

Table 7. *Population Forecast for Da Nang for the Period*

(Thousand persons)

Model		2025	2026	2027	2028	2029	2030
Predicted_Tltds_Model_1	Forecast	2887.8	2960.3	3034.6	3110.7	3188.8	3268.8
	LCL	2829.1	2897.5	2970.2	3044.7	3121.1	3199.5
	UCL	3318.9	3432.4	3518.6	3606.9	3697.4	3709.7

5. Conclusion

The research findings indicate that Da Nang is undergoing a strong urbanization process, as reflected in the increasing proportion of the urban population. This trend not only demonstrates the impacts of economic development, infrastructure investment, and urban spatial expansion, but is also associated with labor migration from rural to urban areas in search of employment opportunities and more stable incomes. In addition, the sex ratio has shown signs of improvement, increasing from approximately 96.69 to 97.91 males per 100 females.

Furthermore, the ARIMAX model results reveal that gender structure has a significant impact on the annual population growth rate. The sex ratio variable has a positive estimated coefficient (23.522) and is statistically significant (Sig. < 0.05), indicating a positive relationship between gender structure and population growth. This finding is consistent with the gender imbalance situation in Da Nang, where an increase in the proportion of males contributes to reducing gender imbalance within the reproductive-age population, thereby increasing the number of births and accelerating population growth [16]. In contrast, the urban–rural population ratio does not appear to have a significant effect on population growth. In addition, the findings regarding the urban–rural population ratio in this study are consistent with several previous studies, which suggest that the relationship between urbanization and population growth is not always clear or statistically significant. In practice, many international studies indicate that the effects of urbanization on demographic variables are often indirect, complex, and highly dependent on the developmental context of each region [17]. In the context of Vietnam, prior studies have also shown that urbanization has occurred rapidly but unevenly across localities, while simultaneously generating various socio-economic consequences such as migration, labor structure transformation, and changes in population distribution [1], [18]. Therefore, the impact of the urban–rural ratio on population size may become "blurred" when other factors, such as migration, employment opportunities, or regional development policies, play a more decisive role.

Based on the above evidence, the fact that the urban–rural ratio variable is statistically insignificant in this study's model should not be considered an abnormal finding, but rather one that is consistent with the general trend observed in prior studies. This further reinforces the argument that the issue should be approached from a multidimensional perspective, rather than expecting a single variable to fully explain population fluctuations.

Forecasting results from the ARIMAX(0,0,1) model suggest that the population of Da Nang during the 2025–2030 period will continue to grow steadily and consistently, creating favorable conditions for socio-economic development, labor force expansion, education, and healthcare services. The population of the city is projected to reach approximately 3,268.8 thousand people by 2030.

Despite the achieved results, the study still has several limitations that should be taken into consideration. First, the sample size remains relatively small, as the data series only covers the period 1999–2024, which may affect the stability and generalizability of the forecasting model. Second, the study only considers a limited number of basic exogenous variables, such as the sex ratio and the urban–rural population ratio, while other important socio-economic factors, including per capita income, employment rate, migration flows, and educational attainment, have not been incorporated into the model. Third, the ARIMAX model employed in this study is primarily based on assumptions of linearity and temporal stability, which may not fully capture nonlinear fluctuations or abnormal shocks occurring in reality. Moreover, the study does not compare the proposed model with alternative forecasting approaches such as VAR models, extended ARIMA models, or machine learning methods; therefore, the relative superiority of the proposed model has not been comprehensively evaluated.

These limitations suggest several directions for future research. Subsequent studies could extend the dataset to include a longer time horizon and incorporate a broader set of socio-economic and policy-related variables to improve model accuracy and explanatory depth. In addition, future research should explore and compare alternative forecasting methods, including both traditional econometric models (e.g., VAR, Bayesian time series models) and modern machine learning approaches, to identify the most suitable techniques for population forecasting. Expanding the analysis to multiple regions or conducting comparative studies across cities would also help enhance the generalizability and practical applicability of the findings.

REFERENCES

[1] General Statistics Office of Vietnam, *Key Findings of the 2024 Mid-term Population and Housing Census as of 00:00 Hours on April 1, 2024*. Hanoi, Vietnam: Statistical Publishing House, 2024.

[2] *Government of Vietnam, Decree on regulations concerning statistical indicators within the national statistical indicator system and the compilation process of GDP and GRDP indicators*, No. 94/2022/ND-CP, 2022.

[3] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control,* 5th ed. Hoboken, NJ, USA: Wiley, 2015.

[4] M. González-Leonardo, M. López-García, and D. R. Paredes, "*Assessing the accuracy of ARIMA models to forecast population at the intraurban level in Greater Mexico City, Mexico*," SocArXiv, 2020.

[5] T. Nyoni, "*Forecasting Total Population in Yemen Using Box-Jenkins ARIMA Models*," MPRA Paper, no. 92433, University Library of Munich, Germany, 2019.

[6] M. Zakria and F. Muhammad, "Forecasting the population of Pakistan using ARIMA models," *Pakistan Journal of Agricultural Sciences*, vol. 46, no. 3, pp. 214–223, 2009.

[7] J. Dai and S. Chen, "The application of ARIMA model in forecasting population data," *Journal of Physics: Conference Series*, vol. 1324, p. 012100, 2019.

[8] M. Gu, "Population prediction of China based on ARIMA-LSTM combined model," *Journal of Sociology and Ethnology*, vol. 5, no. 3, pp. 19–25, 2023.

[9] G. S. Becker, *A Treatise on the Family. Cambridge*, MA, USA: Harvard University Press, 1991.

[10] J. C. Caldwell, *Demographic Transition Theory*. Dordrecht, Netherlands: Springer, 2006.

[11] T. Dyson, *Population and Development: The Demographic Transition*. London, U.K.: Zed Books, 2010.

[12] A. Chipman and E. Morrison, "*The impact of sex ratio and economic status on local birth rates,"* Biology Letters, p. rsbl20130027, 2013.

[13] C. Hirschman, "*Socioeconomic determinants of fertility transition*," in Encyclopedia of Population, P. Demeny and G. McNicoll, Eds. New York, NY, USA: Gale Group, 2003.

[14] J. M. Buresch, D. Medgyesi, J. R. Porter, and Z. M. Hirsch, "Understanding how population change is associated with community sociodemographics and economic outcomes across the United States," *Frontiers in Human Dynamics*, vol. 6, p. 1465218, 2024.

[15] I. Al-Zraigat, A. Al-Omari, and O. Al-Momani, "ARIMA and ARIMAX stochastic models for forecasting live births," *Journal of Statistical Research*, vol. 51, no. 2, pp. 45–56, 2017.

[16] T. L. Y. Ha, "Factors affecting fertility in Danang city," *The University of Danang - Journal of Science and Technology*, vol. 12, no. 73.1, pp. 73–77, 2013.

[17] J. Montalvo, E. Ruiz-Labrador, P. Montoya-Bernabéu, and B. Acosta-Gallo, "Rural–urban gradients and human population dynamics," *Sustainability*, vol. 11, no. 11, p. 3107, 2019.

[18] N. D. T. Nguyen, "The impact of urbanisation on labour resources of Danang city from 2005 to the present," *The University of Danang - Journal of Science and Technology*, vol. 4, no. 101, pp. 20–25, Apr. 2016.