Received: 23 September 2024 | Revised: 18 March 2025 | Accepted: 4 September 2025 | Published online: 23 September 2025

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2025, Vol. 00(00) 1-9
DOI: 10.47852/bonviewJDSIS52024415
BON VIEW PUBLISHING

Building the ARIMA Model for Forecasting the Unemployment Rate by Gender in Vietnam

Quang Phung Duy[1,*], Huyen Linh Nguyen[2], Tra My Dinh Thi[2], Ha Anh Ta Thi[2], Phuong Thuy Nguyen[2], Trung Kien Hoang[2]

[1] *Faculty of Technology and Data Science, Foreign Trade University, Vietnam*

[2] *School of Economics and International Business, Foreign Trade University, Vietnam*

Abstract: Unemployment is a significant macroeconomic issue that affects not only economic development but also the social and psychological well-being of individuals and families. In Vietnam, growing attention has been given to the potential disparity in unemployment rates between men and women, as such differences may reflect structural inequalities in the labor market. This study applies the ARIMA (AutoRegressive Integrated Moving Average) model to forecast gender-specific unemployment rates using official data from 2011 to 2023. The results reveal that although there are fluctuations in unemployment rates for both genders throughout the observed period, the differences are relatively small and not statistically significant. Nonetheless, even minor disparities may suggest the existence of unequal access to employment opportunities, particularly for women in certain regions or sectors. Based on these findings, the report proposes several policy recommendations aimed at fostering a more inclusive labor market. These include expanding access to vocational education and skills development programs for women, implementing gender-sensitive employment policies, and encouraging both public and private sectors to adopt equitable recruitment practices. Furthermore, the enhancement of labor market monitoring systems and the continued use of forecasting tools like ARIMA can support policymakers in anticipating changes and designing timely, data-driven responses. Overall, this study contributes to a deeper understanding of gender-related employment trends in Vietnam and provides practical insights for government agencies, researchers, and policy planners working to promote equal labor market participation and reduce unemployment across all segments of the population.

Keywords: ARIMA, forecast, unemployment rate, gender, macroeconomic

1. Introduction

Research on gender disparities in unemployment and income has highlighted various factors influencing the differences between men and women. Queneau and Sen [1] noted that the unemployment rate gap between men and women in Organization for Economic Cooperation and Development (OECD) countries is due to differences in job-seeking behavior and job distribution. Bičáková [2] further explained that in the European Union (EU), this disparity largely stems from the tendency of women to leave the labor force after childbirth. A study by Heyne and Voßemer [3] revealed that women are less affected by unemployment than men and this difference is attributed not only to financial factors but also to non-financial factors, such as social culture and attachment to the labor market. Ivandić and Lassen [4] find that women experience a 40%–45% higher risk of unemployment and greater earning losses than men in the 2 years following job displacement. In Vietnam, Huyền and Ngân [5] reported that gender income inequality persists; however, it is not as severe compared with that in the region and the world. To mitigate this inequality, Vietnam needs to change the traditional preference for males over females and promote equal job opportunities for both genders.

Research on unemployment rates has attracted the attention of many scholars worldwide, who employ various methods to predict and analyze trends in this index. Mahmudah [6] in Indonesia used time series data from 1986 to 2015, indicating that the ARIMA (AutoRegressive Integrated Moving Average) (0,2,1) model is the optimal forecasting method, with the unemployment rate projected to reach 6.1% by the end of 2016 and 6.0% by the end of 2017. In Greece, Nikolaos et al. [7] emphasized the importance of forecasting unemployment rates in formulating effective fiscal policy strategies, utilizing the Box–Jenkins method to identify the most suitable model and assess forecast accuracy using nonlinear maximum likelihood optimization techniques. Similarly, Adenomon [8] studied the unemployment rates in Nigeria using data from 1974 to 2014, showing high unemployment rates during the period from 2015 to 2018 and proposing that the government needs to create jobs to mitigate negative impacts. Finally, the research conducted by Rublikova and Lubyova [9] in Slovakia applied the ARIMA(0,1,2)(0,1,1)12 + GARCH(1,1) model, demonstrating that this model predicts well with significant variance values. The literature review indicates that ARIMA models and optimization methods are becoming essential tools in the research and forecasting of unemployment rates across various countries. Although general support for gender equality is growing across the MENA region, attitudes remain mixed when it comes to women's participation in the workforce. According to Barometer [10], many people agree in principle with equal rights but support declines when women take on paid work or leadership roles. This reflects deep-rooted social norms that still associate women primarily with domestic responsibilities. The report suggests that increasing women's labor force participation could shift perceptions over time but resistance remains strong in the short term due to traditional beliefs. Donnelly [11] finds that precarious work, such as part-time, poor-quality jobs

*Corresponding author: Quang Phung Duy, Faculty of Technology and Data Science, Foreign Trade University, Vietnam, Email address: quangpd@ftu.edu.vn

and unemployment, is significantly linked to poorer self-rated health for both men and women. The study also reveals that these negative health effects are less pronounced for men in regions or occupations with higher unemployment rates, a pattern not observed among women.

D. Yamacli and S. Yamacli [12] conducted a study to analyze the accuracy of two methods, ARIMA and Artificial Neural Networks, in estimating the unemployment rate in Turkey using monthly data from January 2008 to December 2022. The results showed that the ARIMA (2,1) model was the best fit for forecasting unemployment rates. At the same time, research by Ng et al. [13] compared the performance of ARIMA, SARIMA, and GARCH models in modeling and forecasting unemployment rates during the COVID-19 pandemic across five ASEAN countries, including Malaysia, Singapore, Thailand, the Philippines, and Indonesia, with monthly data from January 2010 to December 2021. The findings indicated that ARIMA and SARIMA models performed more effectively than the GARCH model in modeling and forecasting unemployment rates in the ASEAN-5 countries during the pandemic. Grey-based models have been effectively applied to forecast post-pandemic unemployment in Vietnam [14]. Ning [15] demonstrates that during the COVID-19 pandemic in the United States, women experienced a sharper rise in unemployment compared with men, as shown by deviations from ARIMA-based baseline forecasts.

Applying the ARIMA model to Vietnam's gender-based unemployment data will allow for a more structured analysis of labor market trends, helping policymakers develop targeted strategies to reduce gender disparities and improve employment opportunities for both men and women. This research will not only address a critical gap in understanding gender-specific unemployment patterns but also provide a robust foundation for future economic and social policy development [16].

Unemployment is a notable macroeconomic issue in the present age. According to a report by the International Labor Organization [17], the unemployment rate reached 5.1% in 2023, a modest improvement compared with 5.3% in 2022. However, it is forecasted to exhibit an upward trend in 2024. Unemployment greatly impacts the economy, families, health, and society. Economically, a high unemployment rate can increase poverty [18]. In 2021, Priambodo [19], when analyzing the impact of the unemployment rate on economic growth in Purbalingga Regency, affirmed the existence of a negative relationship between the unemployment rate and economic growth in this area. In terms of families, unemployment leads to poverty, strains relationships, and damages children's development and future careers (McClelland, 2000). Additionally, unemployment affects both physical and mental health. Unemployment and poverty have significant impacts, leading to higher suicide rates across both high-income countries and low- to middle-income countries globally [20]. Unemployment can also contribute to an increased occurrence of depressive symptoms and major depressive disorder, negatively impacting the mental well-being of those without work [21]. Besides, the unemployed tends to visit doctors more, takes more medicines, and spends more days in bed sick [22]. According to a study by Sundar et al. [23], an increase in the unemployment rate can result in the rise of social crimes within society. Unemployment also lowers overall well-being, while poor well-being can also result in unemployment, suggesting that individuals may become caught in a cycle of poor well-being and joblessness [24]. Due to negative impacts in various aspects, the issue of unemployment always receives special attention from the government, policymakers, and researchers. In Vietnam, unemployment is also one of the factors influencing the country's economic development. According to the General Statistics Office [25] in 2023, 1.07 million people in the labor force aged 15 years and above in Vietnam were out of work, showing a decrease of 14.6 thousand people compared with that in 2022. Particularly, there was a difference between the unemployment rates for men and women. The data from the 2022 Employment and Labor Survey Report indicated that the employment rate of women tends to be lower than that of men and the unemployment rate of women tends to be higher. There was an exception in 2022: The unemployment rates decreased and the unemployment rate of men was higher than that of women (2.36% compared with 2.32%). The difference in unemployment rates between men and women reflects many aspects of the country's social, cultural, and labor policies.

Currently, the ARIMA statistical model has been flexibly applied to forecasts in studies. However, building an effective ARIMA model for forecasting the unemployment rate by gender in Vietnam still faces plenty of difficulties. Nguyen et al. [14] forecasted the unemployment rate in Vietnam for 2020–2025 based on data from 2014 to 2019. However, their study had several limitations, as the ARIMA model requires long-term and stable data and the research only used monthly data without considering the impact of quarterly or annual data. Addressing these limitations, this study utilizes quarterly data from 2011 to 2023 and separately analyzes the unemployment rate by gender. Recognizing the urgency of predicting the unemployment rate for sustainable economic development, this study seeks to enhance public awareness of labor market trends and promote gender equality by forecasting the unemployment rate by gender in Vietnam from 2024 to 2025. This article presents a forecasting model based on the ARIMA approach, which can model the unemployment rate across genders. The results indicate minor fluctuations in unemployment rates between men and women. This is a significant topic, as unemployment rates not only reflect economic conditions but also highlight structural differences in the labor market by gender. The key contributions of this study include the following:

1) Developing an optimized ARIMA model to forecast gender-specific unemployment rates in Vietnam for the period 2024–2025, providing a practical approach to labor market analysis.
2) Conducting an in-depth analysis of gender dynamics in the labor market, shedding light on differences in unemployment determinants between men and women.
3) Evaluating the performance of the ARIMA model using statistical criteria such as root mean square error (RMSE), mean absolute error (MAE), and Theil U, ensuring the reliability of the forecasting approach.
4) Proposing socio-economic policies to support labor market development and reduce gender disparities in employment.

2. Data and Methodology

2.1. Data

The research utilizes data on the unemployment rates of men and women in Vietnam, measured as the percentage of unemployed individuals within each gender's total labor force. These data are sourced from the International Labor Organization and published on the World Bank's website [11]. It covers a time series from 2011 to 2023, with figures aggregated by quarter. Data processing and analysis were conducted using Excel and EViews software. The collected data are presented in Table 1.

The dataset covering the period from 2011 to 2023 was selected to ensure a sufficient length for accurate forecasting. However, one limitation that should be acknowledged is the impact of major economic shocks, such as the COVID-19 pandemic or financial crises, which may alter unemployment rate dynamics. Consequently, the forecasting results should be interpreted within a specific economic context, considering additional socio-economic factors that may influence future labor market trends.

Table 1
Unemployment rate by gender data in Vietnam

Quarter	Male	Female	Quarter	Male	Female
2011Q1	1.171	1.521	2014Q2	0.913	1.327
2011Q2	0.887	1.09	2014Q3	1.145	1.526
2011Q3	0.829	0.915	2014Q4	1.163	1.199
2011Q4	0.736	0.836	2015Q1	1.846	1.994
2012Q1	0.902	1.259	2015Q2	1.761	2.076
2012Q2	0.875	1.051	2015Q3	1.653	1.882
2012Q3	0.988	1.154	2015Q4	1.263	1.536
2012Q4	0.929	0.992	2016Q1	1.833	2.477
2013Q1	1.155	1.609	2016Q2	1.917	2.053
2013Q2	1.125	1.442	2016Q3	1.624	1.902
2013Q3	1.294	1.587	2016Q4	1.411	1.514
2013Q4	0.981	1.297	2017Q1	1.998	2.68
2014Q1	1.208	1.537	2017Q2	1.515	1.919
2017Q3	1.84	1.86	2020Q4	2.793	1.482
2017Q4	1.444	1.671	2021Q1	1.755	2.223
2018Q1	1.053	1.422	2021Q2	1.737	1.774
2018Q2	1.124	1.291	2021Q3	2.459	2.623
2018Q3	0.928	1.254	2021Q4	2.602	2.698
2018Q4	0.908	1.154	2022Q1	1.442	1.71
2019Q1	1.493	1.822	2022Q2	1.389	1.722
2019Q2	1.656	1.733	2022Q3	1.457	1.584
2019Q3	1.732	1.695	2022Q4	1.409	1.45
2019Q4	1.508	1.778	2023Q1	1.279	1.789
2020Q1	2.05	1.733	2023Q2	1.552	1.604
2020Q2	2.233	2.252	2023Q3	1.541	1.794
2020Q3	2.729	1.697	2023Q4	1.537	1.708

2.2. Methodology

The ARIMA model was first introduced by Box and Jenkins (1970). This is one of the essential statistical models for time series forecasting. It relies on historical time series data, where the future value is forecasted based on its past movement trends.

Given y_t as the value of the variable at time t, the model is represented as follows:

$$y_t = F(y_{t-1}, y_{t-2}, \ldots, y_0, t)$$

The purpose of this analysis is to elucidate the relationship between the observed y_t values, enabling the prediction of future y_t values. ARIMA is particularly useful for short-term forecasting. The ARIMA model combines both Autoregressive (AR) and Moving Average (MA) models, and it is expressed as follows:

$$\begin{aligned} y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} \\ + \beta_1 \theta_{t-1} + \beta_2 \theta_{t-2} + \ldots + \beta_q \theta_{t-q} + \theta_t \end{aligned}$$

In which y_t is the observed value at time t, are the regression coefficients, and signifies the random error at time t. p and q correspond to the lag orders.

2.2.1. Box–Jenkins method

The Box–Jenkins method is executed through five iterative steps, which provide a comprehensive and systematic framework for constructing the ARIMA model. Therefore, this method ensures the accuracy and performance of the model in forecasting time series.

The Box–Jenkins method consists of the following iterative steps: (i) Checking the stationarity of the time series, (ii) Identifying the tentative model, (iii) Estimation, (iv) Diagnostic checking, and (v) Forecasting, as outlined below:

Step 1: Checking the stationarity of the time series

The data series used in the ARIMA model is assumed to be stationary. Therefore, we need to consider whether these series are stationary or not to predict the unemployment rate by gender in Vietnam when using the ARIMA model. Two common testing methods are the Augmented Dickey–Fuller (ADF) and Phillips–Perron (PP) tests, which econometrics refer to as unit root tests for both the original and differenced data series.

A stationary time series is one in which the variance and expectation remain constant over time. The covariance between two time points does not depend on the specific time points but rather on the lag between them and their respective delays.

Differential (or difference) refers to the discrepancy between the current value and the previous value. Differential analysis is aimed at stabilizing the mean value of a data series, facilitating the transformation of the series into a stationary one.

First-order difference ($I(1)$):

$z(t) = y(t) - y(t - 1)$

Second-order difference ($I(2)$):

$h(t) = z(t) - z(t - 1)$

Step 2: Model identification

The ARIMA model (p, d, q) involves the differencing order (d) of the time series under investigation, the autoregressive order (p), and the moving average order (q). Determining the p and q of the proposed model depends on the plots of the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) of the stationary series. Specifically, the lag p is identified based on the PACF plot and the lag q is determined from the ACF plot.

The selection of p relies on the PACF plot, where the partial autocorrelation coefficients drop abruptly to zero after lag p, while the ACF plot exhibits gradually decreasing autocorrelation coefficients. Similarly, the selection of q is based on the ACF plot, where the autocorrelation coefficients sharply decline after lag q and the PACF plot tapers off.

Step 3: Estimation and Selection of the ARIMA (p, d, q) Model

The parameters of the model are estimated by using EViews software, employing the Conditional Least Squares method and the Newton–Raphson optimization technique. Subsequently, the model is selected based on a comparison of criteria such as adjusted R^2, standard error of regression, AIC (Akaike Information Criterion), and BIC (Bayesian Information Criterion) until the optimal model for forecasting is identified.

Step 4: Model adequacy checking

After determining the parameters of the ARIMA process, it is essential to test whether the model's residuals constitute white noise. We can use the ACF plot or the Breusch–Godfrey test to check for autocorrelation in the residuals. For heteroskedasticity, we can use the White test or the ARCH test. If the chosen model fails these tests, it is necessary to return to the model identification step to select a more appropriate model.

Step 5: Forecasting

Based on the selected ARIMA model, the next step is to determine the point forecasts and the confidence intervals for the forecasts.

3. Results and Estimations

3.1. Unemployment rate in females

3.1.1. Stationarity test

We conduct the stationarity test using the Dickey–Fuller test. The ADF stationarity test results for the male unemployment rate show that the τ statistic ($\tau = -4.702$) is smaller than the critical values at all three significance levels: 1% (−4.148), 5% (−3.500), and 10% (−3.180), with a very small P-value (P-value = 0.0021) (Table 2).

Therefore, it can be concluded that the unemployment rate for females (Female) is stationary. Thus, the ARIMA model for the male unemployment rate has d = 0.

3.1.2. Model identification and estimation

To determine the values of p and q for the model, we rely on the plots of the ACF and the PACF. Observing the PACF plot in Figure 1, we see significant autocorrelation coefficients at lags 1, 2, 8, and 9, leading to the selection of lags $p = 1$, $p = 2$, $p = 8$, and $p = 9$. For the ACF plot, we observe significant partial autocorrelation coefficients at lags 1, 2, 3, and 4, leading to the selection of lags $q = 1$, $q = 2$, $q = 3$, and $q = 4$.

Based on these observations, the ARIMA model can be one of the following forms: ARIMA (1,0,1), ARIMA (1,0,2), ARIMA (1,0,3), ARIMA (1,0,4), ARIMA (2,0,1), ARIMA (2,0,2), ARIMA (2,0,3), ARIMA (2,0,4), ARIMA (8,0,1), ARIMA (8,0,2), ARIMA (8,0,3), ARIMA (8,0,4), ARIMA (9,0,1), ARIMA (9,0,2), ARIMA (9,0,3), or ARIMA (9,0,4).

However, after eliminating models where the intercept term is not statistically significant, we are left with the following models: ARIMA(1,0,1), ARIMA(2,0,1), ARIMA(2,0,2), and ARIMA(8,0,1). To select the best forecasting model, we use an empirical approach by comparing the adjusted R^2, standard error, AIC, and BIC values. The estimation results are presented in Table 3.

Among these criteria, the model with the lowest AIC, BIC, and standard error and the highest adjusted R^2 is selected. The comparison results indicate that the ARIMA (2,0,2) model is the most suitable for the dataset provided by the research group. Table 4 shows the estimation results of the selected model.

3.1.3. Model testing

However, to verify the adequacy of the model, the research group needs to perform additional tests on the residuals to determine whether the model violates any of the regression model assumptions. The White test indicates that the model does not suffer from heteroskedasticity and the Breusch–Godfrey test shows that the model is free from autocorrelation. We can conclude that the model is suitable for forecasting. A detailed summary of the test results is provided in Tables 5 and 6.

3.1.4. Forecasting

The forecasting capability of the model is evaluated using common measurement tools such as the RMSE, MAE, and Theil U coefficient. The Theil U coefficient is a crucial tool for assessing and

Table 2
Augmented Dickey–Fuller test result

		t-Statistic	*P*-Value
Augmented Dickey–Fuller test statistic		−4.702345	0.0021
Test critical value	1% level	−4.148465	
	5% level	−3.500395	
	10% level	−3.179617	

Figure 1
Correlogram of female

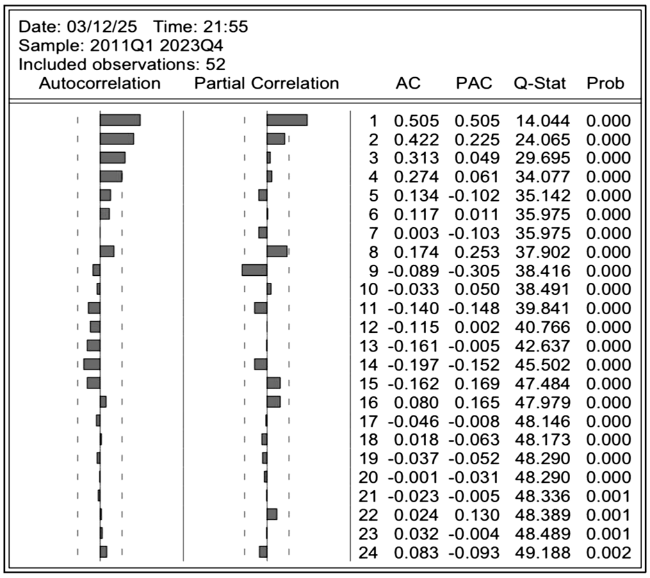


Table 3
Statistical results of a number of standardized ARIMA test models

	Adjusted R^2	SE regression	AIC	BIC
AR (1), MA (1)	0.3103	0.3565	0.8321	0.9457
AR (2), MA (1)	0.2496	0.3689	0.9019	1.0166
AR (1), AR (2), MA (1)	0.3978	0.3305	0.7003	0.8533
AR (1), AR (2), MA (2)	0.4109	0.3269	0.6783	0.8313
AR (2), MA (2)	0.3662	0.3391	0.7329	0.8476
AR (8), MA (1)	0.1345	0.3490	0.7983	0.91997

Table 4
ARIMA model estimate result (2,0,2) for the female series

Variable	Coefficient	Std. error	*t*-Statistic	Prob.
C	1.804044	0.021906	82.35393	0.0000
AR (1)	−0.107187	0.039556	−2.709750	0.0094
AR (2)	0.820139	0.037663	21.77552	0.0000
MA (2)	−0.998427	0.037338	−26.74053	0.0000

Table 5
Breusch–Godfrey serial correlation Language Multiplier (LM) test result

Breusch–Godfrey serial correlation LM test:			
Null hypothesis: no serial correlation at up to 7 lags			
F-Statistic	2.036764	Prob. F (7.39)	0.0747
Ob*R^2	13.38534	Prob. chi-square (7)	0.0633

Table 6
Heteroskedasticity test result

Heteroskedasticity test: White			
Null hypothesis: homoskedasticity			
F-Statistic	0.808947	Prob. *F* (14.35)	0.6544
Ob*R^2	12.22363	Prob. chi-square (14)	0.5883
Scaled explained SS	18.13457	Prob. chi-square (14)	0.2007

comparing the performance of forecasting models; the closer the coefficient is to 0, the higher the forecast model's accuracy.

Calculations in Table 7 show that the forecast results for the four quarters of 2023 are relatively close to the actual values, with the lowest error being 0.1% and the mean error being 0.218%. Additionally, the model's Theil U coefficient (0.089) is less than 1 and close to 0, indicating that the selected ARIMA model is suitable for forecasting and can be applied in practice. The forecast results for 2024 and 2025 are shown in Table 8.

The performance evaluation using RMSE, MAE, and Theil U demonstrates the forecasting effectiveness of the selected ARIMA models. However, as with all econometric models, limitations exist. While ARIMA effectively captures linear patterns, it does not account for potential non-linear relationships in the labor market.

Future studies could further improve predictive accuracy by integrating ARIMA with complementary models, such as neural networks for capturing non-linear trends or economic indicator–based regression models for incorporating external labor market factors. These enhancements could refine labor market forecasting, especially during periods of economic instability.

3.2. Unemployment rate in males

3.2.1. Stationarity test

We conduct the stationarity test using the Dickey–Fuller test. The ADF stationarity test results for the female unemployment rate show that the τ statistic (τ = −3.6653) is smaller than the critical values at the 5% (−3.516) and 10% (−3.188) significance levels, with a very small P-value (P-value = 0.0365) (Table 9).

Therefore, it can be concluded that the unemployment rate for males (Male) is stationary. Thus, the ARIMA model for the female unemployment rate has d = 0.

3.2.2. Model identification and estimation

Observing the PACF plot in Figure 2, we see significant autocorrelation coefficients at lags 1, 5, and 7, leading to the selection of lags $p = 1$, $p = 5$, and $p = 7$. For the ACF plot, we observe significant partial autocorrelation coefficients at lags 1, 2, 3, 4, and 5, leading to the selection of lags $q = 1$, $q = 2$, $q = 3$, $q = 4$, and $q = 5$.

Table 7
The accuracy test of the ARIMA (2,0,2) model's forecasted results

		ARIMA	
Year	**Real value**	**Forecasting value**	**Forecasting error**
Quarter 1 of 2023	1.789	1.772	+1.0%
Quarter 2 of 2023	1.604	1.676	−4.3%
Quarter 3 of 2023	1.794	1.796	−0.1%
Quarter 4 of 2023	1.708	1.713	−0.3%
RMSE		0.303	
MAE		0.218	
Theil U		0.089	

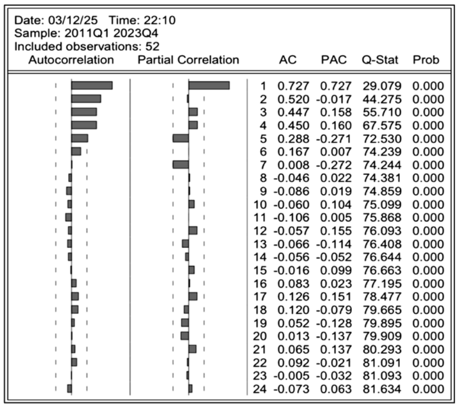
Table 8
Unemployment rate in female forecasted result

Quarter	2024Q1	2024Q2	2024Q3	2024Q4
Forecasting rate	1.835	1.774	1.832	1.776
Quarter	2025Q1	2025Q2	2025Q3	2025Q4
Forecasting rate	1.830	1.778	1.828	1.780

Table 9
Augmented Dickey–Fuller test result

		t-Statistic	*P*-value
Augmented Dickey–Fuller test statistic		−3.665300	0.00356
Test critical value	1% level	−4.180911	
	5% level	−3.515523	
	10% level	−3.188259	

Figure 2
Correlation of male



Based on these observations, the ARIMA model can be one of the following forms: ARIMA (1,0,1), ARIMA (1,0,2), ARIMA (1,0,3), ARIMA (1,0,4), ARIMA (1,0,5), ARIMA (5,0,1), ARIMA (5,0,2), ARIMA (5,0,3), ARIMA (5,0,4), ARIMA (5,0,5), ARIMA (7,0,1), ARIMA (7,0,2), ARIMA (7,0,3), ARIMA (7,0,4), or ARIMA (7,0,5).

However, after eliminating models where the intercept term is not statistically significant, we are left with the following models: ARIMA (1,0,3), ARIMA (1,0,4), ARIMA (5,0,2), and ARIMA (5,0,5). To select the best forecasting model, we use an empirical approach by comparing the adjusted R², standard error, AIC, and BIC values. The estimation results are presented in Table 10.

Table 10
Statistical results of a number of standardized ARIMA test models

	Adjusted R²	**SE regression**	**AIC**	**BIC**
AR (1), MA (2), MA (3)	0.569	0.327	0.675	0.826
AR (1), MA (4)	0.591	0.318	0.603	0.717
AR (5), MA (2)	0.157	0.439	1.257	1.375
AR (1), AR (5), MA (1), MA (2)	0.529	0.329	0.713	0.910
AR (5), MA (5)	0.332	0.392	1.024	1.142

The comparison results indicate that the ARIMA (1,0,4) model is the most suitable for the dataset provided by the research group. Table 11 illustrates the estimation results of the selected model.

3.2.3. Model adequacy testing

The White test indicates that the model does not suffer from heteroskedasticity, while the Breusch–Godfrey test shows that the model is free from autocorrelation. We can conclude that the model is suitable for forecasting. The detailed test results can be found in Tables 12 and 13.

3.2.4. Forecasting

Calculations presented in Table 14 show that the forecast results for the four quarters of 2023 are relatively close to the actual values, with the lowest error being 4.3% and the mean error being 0.240%. Additionally, the Theil U coefficient (0.099) of the model is less than 1 and close to 0.

Therefore, we can conclude that the selected ARIMA model is suitable for forecasting and can be applied in practice. The forecast results for 2024 and 2025 are shown in Table 15.

Table 11
ARIMA model estimate result (1,0,4) for the male series

Variable	Coefficient	Std. error	t-Statistic	Prob.
C	1.501764	0.194869	7.706474	0.0000
AR (1)	0.668785	0.109646	6.099484	0.0000
MA (4)	−0,469212	0.130075	3.607250	0.0007

Table 12
Breusch–Godfrey serial correlation LM test result

Breusch–Godfrey serial correlation LM test:			
Null hypothesis: no serial correlation at up to 7 lags			
F-Statistic	0.270287	Prob. *F* (2.46)	0.7644
Ob*R^2	0.592370	Prob. chi-square (7)	0.7436

Table 13
Heteroskedasticity test result

Heteroskedasticity test: White			
Null hypothesis: homoskedasticity			
F-Statistic	1.969089	Prob. *F* (9.41)	0.0747
Ob*R^2	15.39142	Prob. chi-square (9)	0.0633
Scaled explained SS	12.17355	Prob. chi-square (9)	0.2037

Table 14
The accuracy test of the ARIMA (1,0,4) model's forecasted results

		ARIMA	
Year	Real value	Forecasting value	Forecasting error
Quarter 1 of 2023	1.279	1.224	−4.3%
Quarter 2 of 2023	1.552	1.336	−13.9%
Quarter 3 of 2023	1.541	1.432	−7,1%
Quarter 4 of 2023	1.537	1.451	−5.6%
RMSE	0.308		
MAE	0.240		
Theil U	0.099		

Table 15
Unemployment rate in male forecasted result

Quarter	2024Q1	2024Q2	2024Q3	2024Q4
Forecasting rate	1.551	1.636	1.643	1.636
Quarter	2025Q1	2025Q2	2025Q3	2025Q4
Forecasting rate	1.592	1.562	1.542	1.529

4.Results and Discussion

4.1. Results

The authors built an ARIMA model and thereby made forecasts of the unemployment rate in Vietnam by gender in four quarters of 2024 and 2025. The ARIMA (1,0,4) model and ARIMA (2,0,2) model were used to forecast the unemployment rates of men and women in each quarter of 2024 and 2025 and obtained the results in the following order: 1.551%, 1.636%, 1.643%, 1.636%, 1.592%, 1.562%, 1.542%, 1.529% (men), 1.835%, 1.774%, 1.832%, 1.776%, 1.830%, 1.778%, 1.828%, and 1.780% (women). Accordingly, we see that the forecasted unemployment rate for men in 2024 shows a slight increase compared with that in 2023 but a gradual decrease is observed in 2025. Meanwhile, the unemployment rate for women remains relatively stable throughout 2024 and 2025 with minor fluctuations. However, the general trend suggests that women's unemployment remains higher than men's in all quarters, indicating persistent gender disparities in employment opportunities.

It can be observed that while the unemployment rate for both genders fluctuates within a narrow range, the gap between male and female unemployment persists. Although the absolute differences are relatively small, the consistently higher female unemployment rate suggests structural labor market challenges for women. This does not necessarily indicate low gender inequality, as female workers are predominantly concentrated in vulnerable jobs, particularly domestic and informal work. Vietnamese women face a disproportionate "double burden" due to both paid and unpaid labor responsibilities, resulting in lower incomes and fewer opportunities for career advancement [16].

Furthermore, contrary to common expectations, the labor force participation rate for men (62.6%) is lower than that for women (74.7%) in the first quarter of 2024. This deviation may be influenced by gender-based employment preferences, household responsibilities, or industry-specific labor demands. Despite their higher labor force participation, women still experience significant economic disparities compared with men in terms of income levels, job quality, and leadership representation.

The findings align with previous labor market studies but highlight certain contradictions that warrant further investigation. In particular, the relatively higher unemployment rate for men in specific quarters challenges conventional assumptions that women face greater employment difficulties. This discrepancy suggests that male-dominated industries, such as construction and manufacturing, may be more vulnerable to economic fluctuations, whereas women's employment is more affected by social policies, caregiving responsibilities, and job stability in sectors such as education and healthcare.

The variation in selected ARIMA models—ARIMA (2,0,2) for women and ARIMA (1,0,4) for men—reflects distinct gender-based labor market dynamics. The higher-order moving average terms in the female unemployment model may indicate external social and policy-driven influences, whereas the autoregressive terms in the male model suggest a stronger dependence on economic cycles. This aligns with previous research indicating that male employment is more cyclical, while female employment is shaped by long-term social and institutional factors.

Moreover, while RMSE, MAE, and Theil U values confirm the reliability of the selected ARIMA models, further robustness checks could strengthen the analysis. Future research could incorporate sensitivity analyses to assess the impact of external shocks—such as economic recessions or changes in labor policies—on model stability. This would provide a more comprehensive understanding of how Vietnam's labor market responds to macroeconomic and social changes.

4.2. Policy implications

Unemployment is a major issue related to people's livelihoods, social consequences, and the economic growth of the country. Based on the forecast results and the current labor market situation in Vietnam, the authors suggest several policies for the government:

Labor Market Development: Actively participate in international integration and global economic activities. Develop and implement appropriate human resource development strategies to align with economic globalization. Enhance policies to guide the labor market in a sustainable direction, ensuring that minimum wages provide a living income for workers and enforcing labor rights protections through inspections and regulatory measures.

Workforce Quality Enhancement: Improve education and vocational training to match labor market demands. Strengthen collaboration between educational institutions and businesses to align training programs with real-world needs. Additionally, integrate administrative measures to instill an industrial work ethic in workers, ensuring skill proficiency and professionalism.

Social Security System Development: Establish a comprehensive social security system encompassing financial support, self-employment opportunities, career transitions, job counseling, and labor market information updates. Provide effective support programs for the unemployed, including retraining initiatives and job search assistance.

Gender Equality in Employment: Policies should address gender disparities in the labor market. Encouraging women to take on leadership roles in both public and private sectors can help reduce gender inequality. Providing leadership training and skill development programs for women will enhance their career advancement opportunities. Additionally, improving job quality and workplace conditions for female workers, including implementing family-friendly policies such as childcare support, maternity leave, and flexible work arrangements, will enable better work–life balance.

These policies directly respond to the forecasted unemployment trends. The relatively stable unemployment rate suggests that current labor market conditions are favorable, but targeted interventions can further support vulnerable groups, particularly women. Ensuring sustainable employment and workforce development is crucial for Vietnam's continued economic growth and social stability.

5. Conclusions

This study applied ARIMA models to forecast Vietnam's unemployment rate by gender, providing insights into labor market trends for 2024 and 2025. The findings indicate that the unemployment rates for both men and women remain relatively stable, with only minor fluctuations across quarters. However, while the absolute differences between male and female unemployment rates are small, women consistently experience higher unemployment rates than men in each quarter. This suggests that, despite having a high-labor force participation rate, women continue to face labor market disadvantages due to systemic inequalities, including lower job quality, wage gaps, and underrepresentation in decision-making positions.

While various alternative forecasting methods exist, including SARIMA, GARCH, and machine learning models, ARIMA was chosen due to its proven effectiveness in short-term economic forecasting and its interpretability. Unlike machine learning models, which require extensive data preprocessing and may be difficult to interpret, ARIMA provides clear, statistically grounded predictions based on past unemployment trends. Additionally, SARIMA is more beneficial for seasonally dependent time series, which does not apply to Vietnam's unemployment data, while GARCH is better suited for modeling financial volatility rather than labor market trends.

However, future research could compare the performance of ARIMA with alternative methods, such as SARIMA, machine learning models, or hybrid approaches, to assess their relative forecasting accuracy. Such studies could help determine whether combining traditional time series models with advanced forecasting techniques might enhance predictive performance while maintaining interpretability.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data are available from the corresponding author upon reasonable request.

Author Contribution Statement

Quang Phung Duy: Conceptualization, methodology, writing – review & editing, supervision, and project administration. **Huyen Linh Nguyen:** Methodology, formal analysis, investigation, data curation, writing – original draft, writing—review & editing, and project administration. **Tra My Dinh Thi:** Software, validation, formal analysis, data curation, writing – original draft, writing – review & editing, and project administration. **Ha Anh Ta Thi:** Investigation, writing – original draft, writing – review & editing, visualization, and project administration. **Phuong Thuy Nguyen:** investigation, resources, data curation, writing – original draft, writing – review & editing, and project administration. Trung Kien Hoang: Investigation, writing – original draft, writing – review & editing, and project administration.

References

[1] Queneau, H., & Sen, A. (2008). Evidence on the dynamics of unemployment by gender. *Applied Economics*, *40*(16), 2099–2108. https://doi.org/10.1080/00036840600949330

[2] Bičáková, A. (2016). Gender unemployment gaps in the EU: Blame the family. *IZA Journal of European Labor Studies*, *5*, 1–31. https://doi.org/10.1186/s40174-016-0072-3

[3] Heyne, S., & Voßemer, J. (2023). Gender, unemployment, and subjective well-being: Why do women suffer less from unemployment than men?. *European Sociological Review*, *39*(2), 301–316. https://doi.org/10.1093/esr/jcac030

[4] Ivandić, R., & Lassen, A. S. (2023). Gender gaps from labor market shocks. *Labour economics*, *83*, 102394. https://doi.org/10.1016/j.labeco.2023.102394

[5] Huyền, N. T. T., & Ngân, N. T. (2018). Chênh lệch thu nhập theo giới tính: lý thuyết và thực trạng tại Việt Nam. *TNU Journal of Science and Technology*, *191*(15), 93–98.

[6] Mahmudah, U. (2017). Predicting unemployment rates in Indonesia. *Economic Journal of Emerging Markets*, *9*(1), 20. http://dx.doi.org/10.20885/ejem.vol9.iss1.art3

[7] Nikolaos, D., Stergios, A., Tasos, S., & Ioannis, S. (2018). Forecasting unemployment rates in Greece. *International Journal of Sciences: Basic and Applied Research*, *37*(1), 43–55.

[8] Adenomon, M. O. (2017). Modelling and forecasting unemployment rates in Nigeria using ARIMA model. *FUW Trends in Science and Technology Journal*, *2*(1B), 525–531.

[9] Rublikova, E., & Lubyova, M. (2013). Estimating ARIMA-ARCH model rate of unemployment in Slovakia. *Forecast. Pap. Progn. Pr*, *5*, 275–289.

[10] Arab Barometer (2024). Gender attitudes and trends in MENA: The effects of working women. *Arab Barometer*. Available at: https://www.arabbarometer.org/wp-content/uploads/AB8-Gender-Report- EN.pdf

[11] Donnelly, R. (2021). Precarious work and heath: Do occupation-and state-specific unemployment rates matter for women and for men? *SSM-Population Health*, *16*, 100967. https://doi.org/10.1016/j.ssmph.2021.100967

[12] Yamacli, D. S., & Yamacli, S. (2023). Estimation of the unemployment rate in Turkey: A comparison of the ARIMA and machine learning models including Covid-19 pandemic periods. *Heliyon*, *9*(1). https://doi.org/10.1016/j.heliyon.2023.e12796

[13] Ng, K. Y., Zainal, Z., & Samsudin, S. (2023). Comparative performance of ARIMA SARIMA and GARCH models in modelling and forecasting unemployment among ASEAN-5 countries. *International Journal of Business & Society*, *24*(3). https://doi.org/10.33736/ijbs.6393.2023

[14] Nguyen, P. H., Tsai, J. F., Kayral, I. E., & Lin, M. H. (2021). Unemployment rates forecasting with grey-based models in the post-COVID-19 period: A case study from Vietnam. *Sustainability*, *13*(14), 7879. https://doi.org/10.3390/su13147879

[15] Ning, K. (2024). The influence of gender disparities on the unemployment rate in the United States amid the COVID-19 pandemic. *Advances in Economics, Management and Political Sciences*, *57*, 268–275. https://doi.org/10.54254/2754-1169/57/20230769

[16] Vo, D. H., & Ho, C. M. (2023). Does gender and education of the households' heads matter for wealth accumulation in Vietnam? Evidence from a recent decade. *Heliyon*, *9*(12). https://doi.org/10.1016/j.heliyon.2023.e22836

[17] International Labour Organization (ILO). (2024). Global unemployment rate set to increase in 2024 while growing social disparities persist. Available at: https://www.ilo.org/resource/news/global-unemployment-rate-set-increase-2024-while-growing-social

[18] Yasmin, F., Shaheen, R., Yasin, A., & Naseer, S. (2020). An analysis of causal relationship between economic growth and unemployment: Evidence from Pakistan. *Annals of Social Sciences and Perspective*, *1*(1), 09–17. https://doi.org/10.52700/assap.v1i1.14

[19] Priambodo, A. (2021). The impact of unemployment and poverty on economic growth and the human development index (HDI). *Perwira International Journal of Economics & Business*, *1*(1), 29–36.

[20] Sinyor, M., Silverman, M., Pirkis, J., & Hawton, K. (2024). The effect of economic downturn, financial hardship, unemployment, and relevant government responses on suicide. *The Lancet Public Health*, *9*(10), e802–e806. https://doi.org/10.1016/S2468-2667(24)00152-X

[21] Amiri, S. (2022). Unemployment associated with major depression disorder and depressive symptoms: A systematic review and meta-analysis. *International Journal of Occupational Safety and Ergonomics*, *28*(4), 2080–2092. https://doi.org/10.1080/10803548.2021.1954793

[22] Linn, M. W., Sandifer, R., & Stein, S. (1985). Effects of unemployment on mental and physical health. *American Journal of Public Health*, *75*(5), 502–506. https://doi.org/10.2105/AJPH.75.5.502

[23] Sundar, S., Tripathi, A., & Naresh, R. (2018). Does unemployment induce crime in society? A mathematical study. *American Journal of Applied Mathematics and Statistics*, *6*(2), 44–53.

[24] Gedikli, C., Miraglia, M., Connolly, S., Bryan, M., & Watson, D. (2023). The relationship between unemployment and wellbeing: An updated meta-analysis of longitudinal evidence. *European Journal of Work and Organizational Psychology*, *32*(1), 128–144. https://doi.org/10.1080/1359432X.2022.2106855

[25] General Statistics Office of Vietnam. (2025). *Báo cáo điều tra lao động việc làm năm 2023 [Report on labour force survey 2023]*. Statistical Publishing House. https://www.nso.gov.vn/default/2025/03/bao-cao-dieu-tra-lao-dong-viec-lam-nam-2023/

How to Cite: Duy, Q. P., Nguyen H. L., Thi, T. M. D., Thi, H. A. T., Nguyen P. T., & Hoang, T. K. (2025). Building the ARIMA Model for Forecasting the Unemployment Rate by Gender in Vietnam. *Journal of Data Science and Intelligent Systems*. https://doi.org/10.47852/bonviewJDSIS52024415