# Research on the Unemployment Rate in Texas based on ARIMA Model

Xiuyuan ShangGuan*

Department of mathematics, Nanjing University, Nanjing, China

*Corresponding author: 201010013525@stu.swmu.edu.cn

**Abstract.** Since the outbreak of the COVID-19 pandemic in 2020, the issue of unemployment has garnered widespread attention from the government, the public, employers and various other stakeholders. In this context, accurate prediction of the unemployment rate has become particularly crucial. This study builds upon previous research and focuses on the unemployment rate data in Texas from January 1976 to September 2023. Firstly, the data undergoes first-order differencing to transform it into a stable time series. Subsequently, the author constructs ARIMA models with different parameters and select appropriate evaluation metrics to obtain more accurate forecasting results. Finally, in conjunction with the realities of the COVID-19 pandemic, a comprehensive analysis is conducted on both the predictive outcomes and the original data. Analysis of the raw data shows that the impact of the COVID-19 pandemic on the unemployment rate is undeniable. The forecasting results indicate an upward trend in the unemployment rate for the state of Texas. Based on these analyses, this paper presents some policy recommendations to the government in addressing the unemployment issue.

**Keywords:** Unemployment rate; Texas; ARIMA model.

## 1. Introduction

The unemployment rate served as a pivotal gauge reflecting the supply and demand dynamics within the labor market and stood as a primary target for macroeconomic regulation across nations. According to Levine et al [1], the primary determinant of the unemployment rate, from a public policy perspective, was the pace of economic growth. Notably, the eruption of the COVID-19 pandemic induced economic stagnation, resulting in significant fluctuations in the unemployment rate. Findings cited by Ahmad et al. suggested an anticipated rise in the unemployment rate in forthcoming years due to the repercussions of the coronavirus. It was projected that a minimum of 5 years would be necessary to alleviate the impact of COVID-19 in these nations [2].

The causes behind the escalation of the unemployment rate were multifaceted. For instance, Marika Karanassou and others posited that oil price shocks and capital accumulation might contribute to the rise in unemployment rate [3]. The impact mechanisms of these factors on the unemployment rate were highly intricate. To more directly address the influence of these factors, redirecting research focus towards trends in data variation became paramount. This approach could elucidate the underlying principles of unemployment rate growth and facilitate the utilization of computational tools for forecasting its developmental trajectory. In essence, this paper aimed to offer guidance for diverse governmental decision-making.

The essay revolved around forecasting the future unemployment rate in Texas. The state of Texas held a pivotal position in the U.S. economy, boasting a GDP of $13.974 trillion in 2012, marking a year-on-year growth of 4.8%, the second-highest growth rate nationwide. It contributed 9% to the total U.S. economic output, ranking second among the 50 states and earning accolades as a "pacesetter" in the American economic recovery. Texas had consistently maintained an unemployment rate below 10%, drawing significant attention as it remained below the national average [4]. The choice of SPSS was guided by its widespread usage in both academic and business spheres, establishing it as the predominant software in its category. Furthermore, SPSS was a versatile tool capable of facilitating various analyses, data transformations, and output formats, ensuring it effectively meets the demands of our research [5]. Based on the findings of this study, we could extrapolate the general economic trends in Texas. Post-pandemic, when the government sought to rejuvenate the economy, this paper was poised to provide policy recommendations at the governmental level.

In the realm of unemployment rate prediction, Christos Katris et al. employed fully connected feedforward neural networks with support for multivariate adaptive regression splines and vector regression, as well as time series analysis, to forecast unemployment rates in Mediterranean countries [6]. In our exploration, we aimed to determine whether ARIMA models with GARCH errors could yield proper results. To ensure that the chosen models did not exhibit "explosive" behavior or escalating variance over time, Chakraborty et al. presented results on the asymptotic stationarity of their proposed hybrid approach, utilizing Markov chains and nonlinear time series analysis techniques [7]. Finally, Michał Gostkowski et al. described, developed, and compared various predictive methods, including the naive method, regression model, ARIMA, Winters model, and Holt model, using data collected by the Central Statistical Office [8]. Numerous methods and models are currently available for predicting unemployment rates, with time series models being regarded as the most reliable approach.

In summary, Texas stood as a highly representative state, and predicting the unemployment rate in Texas held significant importance for society, the economy, and even employers [9, 10]. Building upon previous research, this paper endeavored to construct ARIMA models with varying parameters to obtain more meaningful predictive results. Additionally, considering real-world factors such as the pandemic's impact on unemployment rates, the study aimed to provide recommendations for the government.

## 2. Methods

### 2.1. Data Source

The data for this literature was collected from the Fred website, and the unemployment rate data for Texas from January 1976 to September 2023 were selected for a total of 573 observations.

### 2.2. Variable Selection

Since the unemployment rate is a long-term statistic to follow, the whole statistic is concluded in the research for different aspects to research. The complete and whole statistics are there for the research for time series and segmented time series, figuring out the main rules of how these trends were changing, which are plotted below in Fig 1.

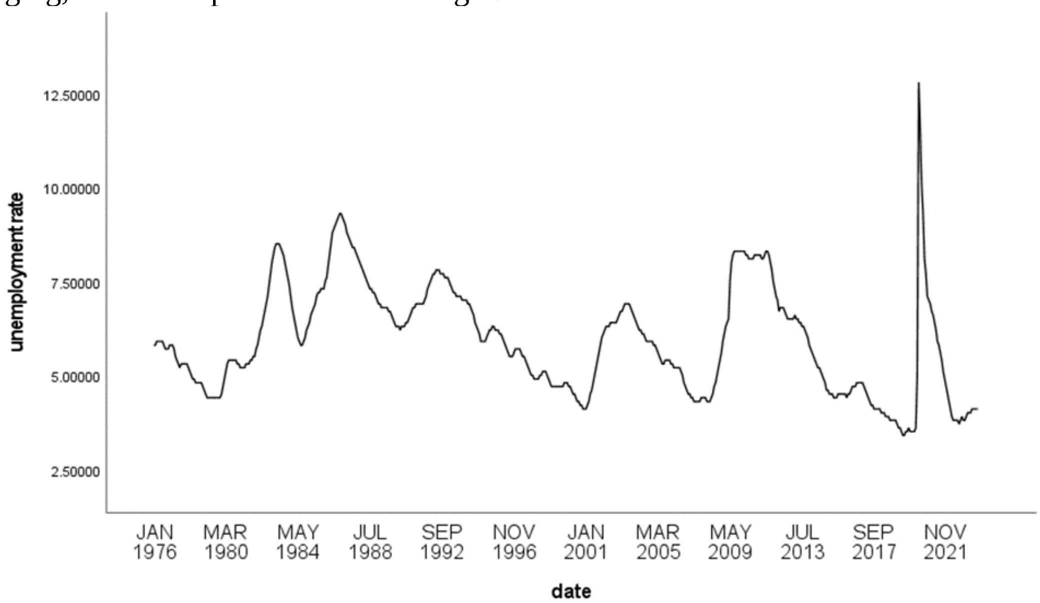


**Fig. 1** Unemployment rates in Texas

The unemployment rate is a major indicator which reflects the degree of labor resource utilization in a country or region. In general, an upward trend in unemployment means that more labor resources are not being used effectively, leading to a rise in the number of unemployed, and with a consequent decrease in the overall demand of society, thus undermining the momentum of economic growth. Consequently, governments worldwide have consistently regarded the unemployment rate as a crucial metric for assessing the macroeconomic condition and the prosperity of the labor market. This assessment serves as a foundation for formulating or adjusting relevant macroeconomic and employment policies. The following chart illustrates the unemployment rate in Texas from January 1976 to September 2023.

Examining the unemployment rate curve yields some straightforward conclusions. Initially, fluctuations in the unemployment rate exhibit apparent periodicity, yet the durations of these cycles vary. Furthermore, unemployment rate data frequently manifest fluctuations but remain stable within the range of 3% to 13%. Lastly, in 2020, there was a significant surge in unemployment, reaching a historic peak of 12.8% in April—marking the first instance of surpassing the 10% threshold. Considering a contextual analysis, this temporal surge aligns with the peak of the COVID-19 pandemic outbreak.

### 2.3. Data Processing

The autocorrelation function (ACF) plot in Fig 2 indicates a pronounced autocorrelation in the original time series data, making direct time series forecasting challenging. Consequently, performing a first-order differencing on the original data is undertaken to further mitigate the robust autocorrelation among the data points.

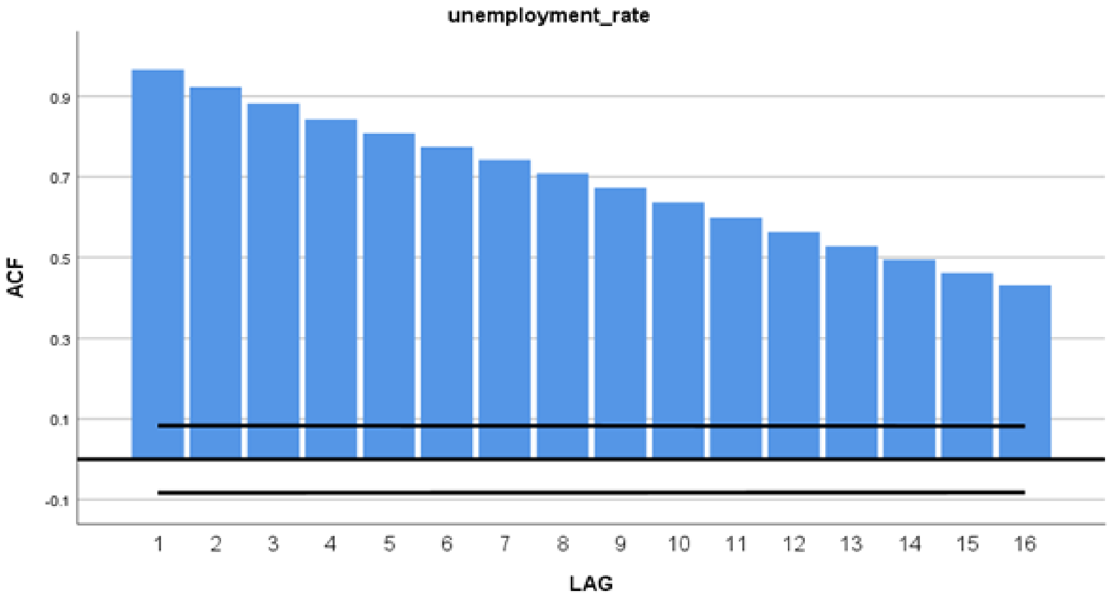


**Fig. 2** The ACF plot of measured data

After applying first-order differencing to the original data and replotting the autocorrelation function (ACF) in Fig 3 and partial autocorrelation function (PACF) in Fig 4, the aim is to assess whether the differenced data exhibits smoothing characteristics. From Fig 2, it is evident that the majority of the autocorrelation coefficients fall within the expected range, displaying a generally decaying pattern. Similarly, the PACF plot yields the same conclusion. Consequently, it can be inferred that, following the first-order differencing, the data can be considered as a stationary time series. Subsequently, appropriate time series analysis models can be employed to analyze the data and forecast future trends.

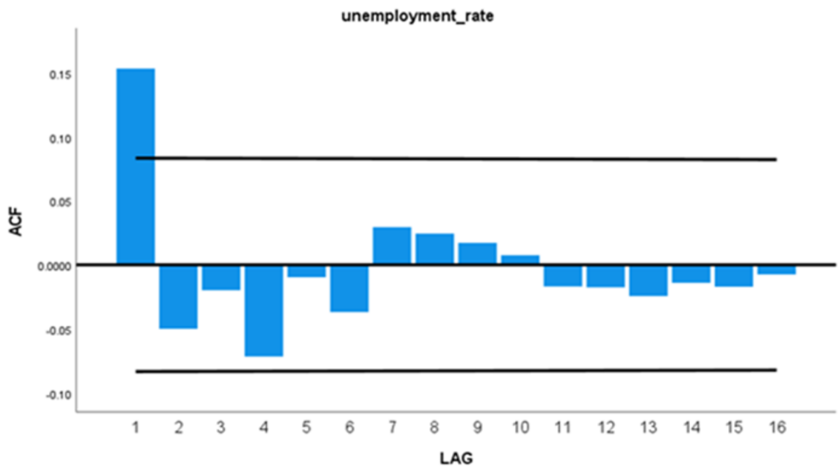


**Fig. 3** The ACF plot of the first-order differenced Measured data

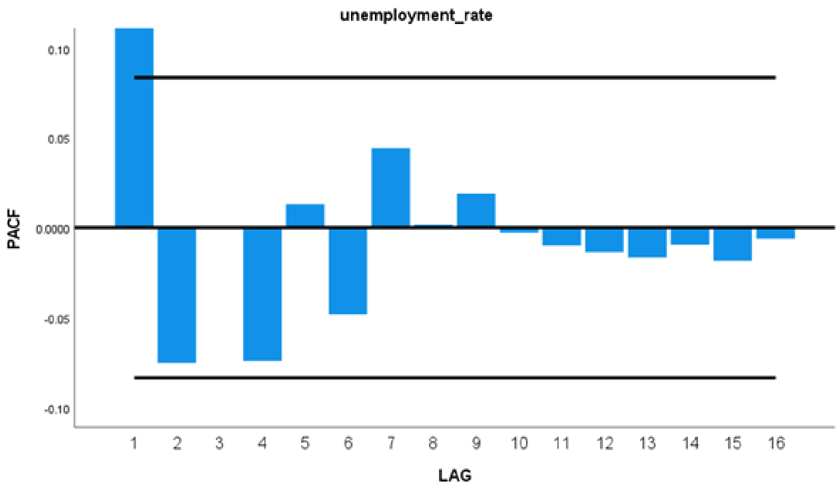


**Fig. 4** The PACF plot of the first-order differenced Measured data

## 2.4. Method Selection

An autoregressive integrated moving average (ARIMA) model is a statistical model used to analyze and forecast time series data. It combines autoregressive (AR) and moving average (MA) methods, combined with differential (ensemble) steps, to convert non-stationary time series data into stationary form. ARIMA is adept at capturing trends and periodicity in time series data. Therefore, this study utilizes the ARIMA model to construct a predictive model for the unemployment rate in the state of Texas.

## 2.5. Method Evaluation

By comparing the original data with the fitted data obtained from the ARIMA model, it is crucial to select appropriate evaluation metrics to determine the accuracy and performance of the model. The Bayesian Information Criterion (BIC) assesses the trade-off between model parameter complexity and performance, aiding in preventing overfitting. On the other hand, Root Mean Square Error (RMSE) is effective in measuring the precision of model fitting. Therefore, this study employs RMSE and BIC as indicators to evaluate the relative merits of each model in Table 1.

**Table 1.** Model evaluation

| Model | RMSE | BIC |
|---|---|---|
| ARIMA(1,1,4)(1,0,1) | 0.385 | -2.020 |
| ARIMA(1,1,3)(1,0,1) | 0.362 | -1.945 |
| ARIMA(1,1,0)(1,0,1) | 0.360 | -1.965 |
| ARIMA(1,0,1) | 0.363 | -1.983 |

The computation of BIC involves the negative part of the log-likelihood function and the number of model parameters, where the negative part of the log-likelihood function typically dominates, resulting in negative values. This phenomenon is normal. According to multiple calculations of the ARIMA model, results as shown in Table 1, it is evident that the ARIMA(1,1,3) model exhibits the highest accuracy. However, considering its higher complexity and the marginal improvement in accuracy compared to the simpler ARIMA(1,0,1) model, the latter is deemed the optimal choice in the current context.

## 3. Results and Discussion

### 3.1. Forecasting

The forecasted results from the ARIMA model indicate an upward trend in future unemployment rates. As computed through SPSS, the predicted unemployment rates for the next 12 months are documented in Table 2. Fig 5 illustrates the model-predicted unemployment rate curve. In Fig 5, the red solid line represents Measured data, the blue thin line represents Fit data, and the purple dashed lines flanking the fitted data indicate the upper and lower confidence interval thresholds (UCL, LCL). The blue bold line represents the Forecast data generated by the ARIMA model.

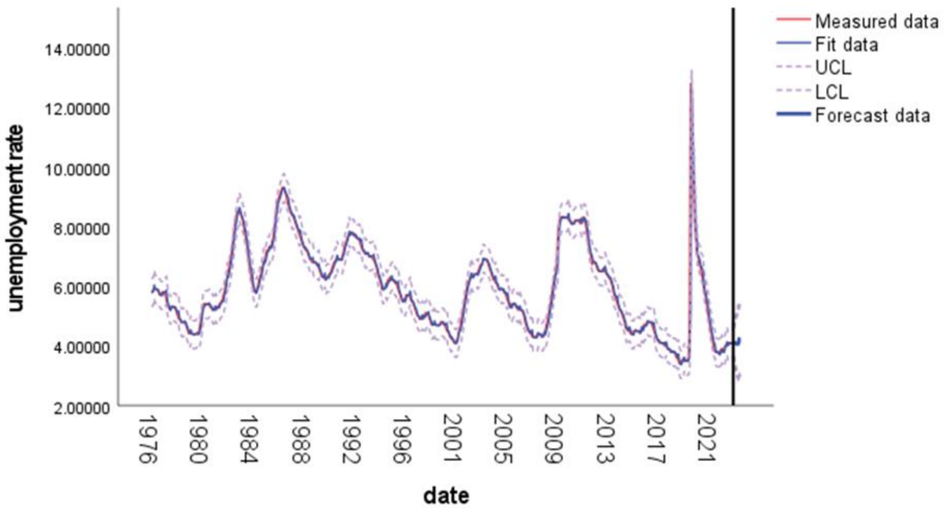


**Fig. 5** Forecasting with model

**Table 2.** Predicting with model

| Time | Point Forecast | Lo 80 | Hi 80 | Lo 95 | Hi 95 |
|---|---|---|---|---|---|
| Oct 2023 | 4.167 | 3.707 | 4.626 | 3.463 | 4.870 |
| Nov 2023 | 4.243 | 3.543 | 4.944 | 3.171 | 5.315 |
| Dec 2023 | 4.316 | 3.453 | 5.180 | 2.994 | 5.638 |
| Jan 2024 | 4.386 | 3.397 | 5.375 | 2.872 | 5.900 |
| Feb 2024 | 4.453 | 3.362 | 5.544 | 2.782 | 6.123 |
| Mar 2024 | 4.516 | 3.340 | 5.693 | 2.715 | 6.318 |
| Apr 2024 | 4.577 | 3.328 | 5.827 | 2.665 | 6.490 |
| May 2024 | 4.635 | 3.323 | 5.948 | 2.627 | 6.644 |
| Jun-2024 | 4.691 | 3.324 | 6.058 | 2.598 | 6.784 |
| Jul-2024 | 4.744 | 3.329 | 6.159 | 2.577 | 6.911 |
| Aug-2024 | 4.795 | 3.337 | 6.253 | 2.563 | 7.026 |
| Sep-2024 | 4.843 | 3.347 | 6.339 | 2.553 | 7.133 |

### 3.2. Residuals

The initial plot in Fig 6 indeed suggests the emergence of white noise alongside an anomalous value around the year 2020.

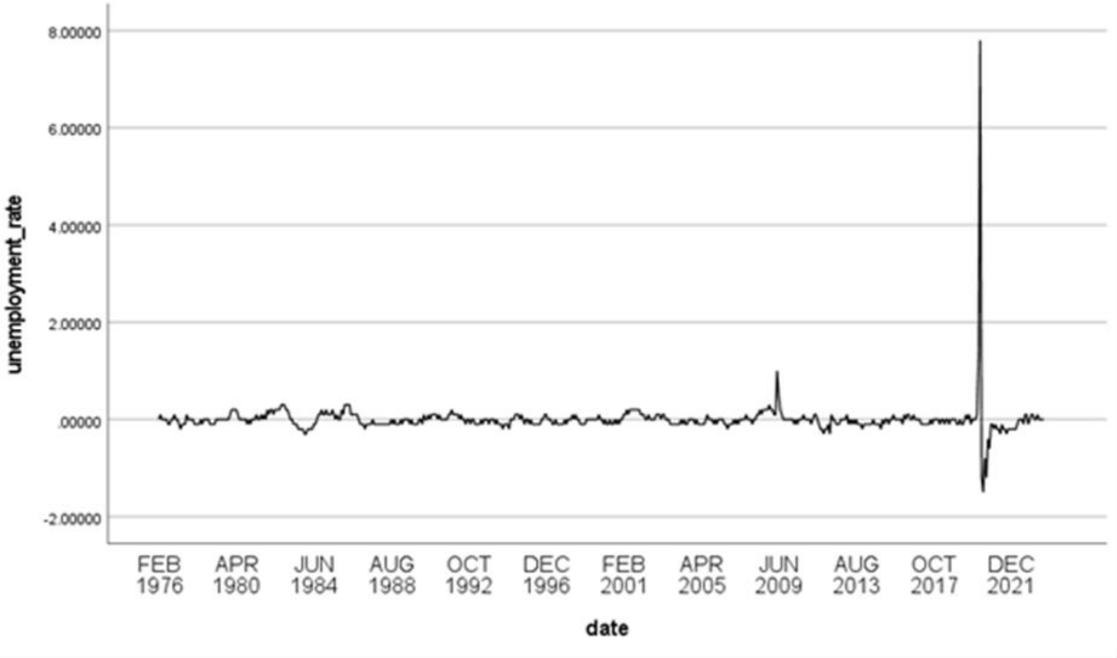


**Fig. 6** The first-order differenced sequence diagram

The autocorrelation function (ACF) plot in Fig 7 indicates that none of the autocorrelations surpass the blue critical values, suggesting that the series appears to resemble white noise. Subsequently, the third plot illustrates that the series is not normally distributed. However, the QQ plot in Fig 8, despite appearing as a straight line, confirms that the series is indeed normally distributed. According to the ACF graph, no autocorrelation values exceed the critical threshold, indicating a white noise nature in the original data. The majority of the data in Fig 8 aligns along a straight line, affirming the approximate fulfillment of normal distribution for the series.

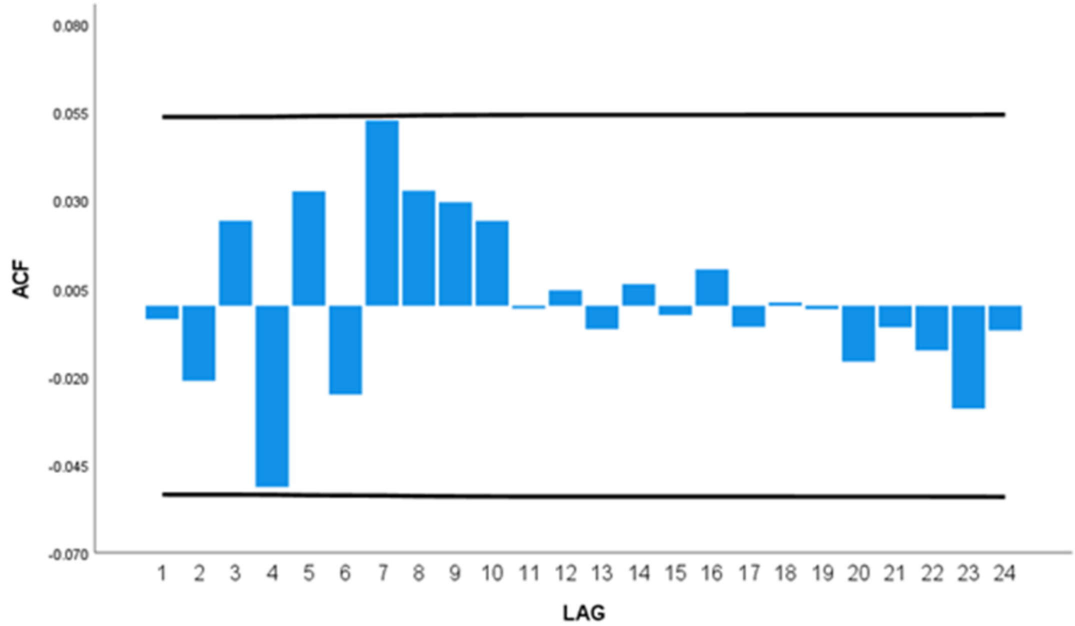


**Fig. 7** The ACF plot of the fit data

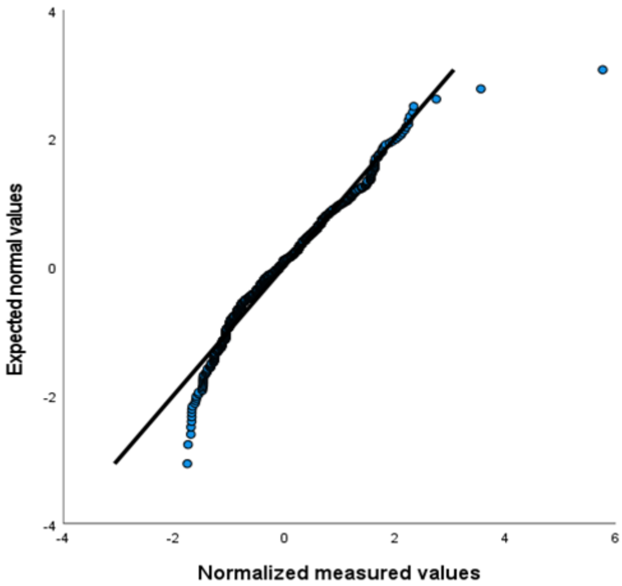


**Fig. 8** QQ plot

As per the autocorrelation function (ACF) plot, none of the autocorrelation values exceed the critical threshold, indicating that the original data exhibits characteristics of white noise. The data in Figure 8 predominantly follows a linear distribution, leading to the conclusion that the series roughly conforms to a normal distribution. The obtained p-value exceeds 0.05, indicating that the initial hypothesis remains valid. There is no autocorrelation observed in the series. Consequently, the predicted data appears to exhibit characteristics similar to white noise. The outcome is deemed satisfactory.

However, there is a crucial point to note: between 2019 and 2022, the unemployment rate in Texas exhibited significant fluctuations. When utilizing time series forecasting for future data, it's common practice to eliminate outliers to enhance the accuracy of the predictive model. Yet, in practical terms, the adverse effects of COVID-19 on markets and the economy are expected to persist for an extended period, both in the United States and globally. Therefore, the results obtained in this study may differ somewhat from the actual future trends in unemployment rates.

### 3.3. Critical Thinking

Despite having obtained forecasted data for future unemployment rates in Texas, the impact of COVID-19 cannot be ignored. Through data analysis and visualization in a box plot in Fig 9, some outliers can be observed. It is noticeable that these outliers predominantly occur after March 2020, coinciding with the onset of the COVID-19 pandemic.

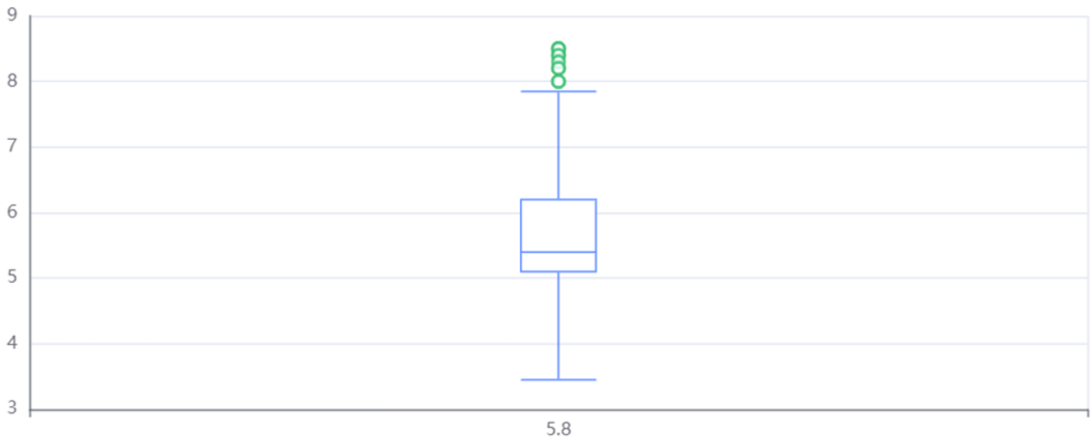


**Fig. 9** Boxplot of unemployment rate

## 4. Conclusion

Through this study, the analysis leads to some conclusions: Firstly, the pandemic has had a dramatically impact on the economy and livelihoods, resulting in a substantial increase in the unemployment rate. While the rate has shown a noticeable decrease after the pandemic, the consequences remain irreparable. Notably, according to ARIMA forecasts, the unemployment rate is expected to rise again in the future.

Therefore, to prevent a substantial surge in unemployment, which could lead to societal panic and reduce the number of unemployed individuals, government departments can consider taking the following measures when formulating policies:

Invest in vocational training and skill development to ensure the workforce adapts to market demands. Provide subsidies or scholarships to encourage individuals to acquire new skills, ensuring that training aligns with market needs. Implement policies to promote entrepreneurship, providing loans and training programs for entrepreneurs to create more job opportunities. Strengthen investment in infrastructure projects such as roads, bridges, airports, etc., to stimulate economic growth and create jobs in construction and maintenance.

Last but not least, the government should boost public confidence and enhance the overall economic and social resilience to disruptions and risks, swiftly moving beyond the shadows cast by the pandemic.

## References

[1] Linda Levine. Economic growth and the unemployment rate. Congressional Research Service, 2013.

[2] Ahmad M, Khan Y A, Jiang C, Kazmi S J H, Abbas S Z. The impact of COVID-19 on unemployment rate: an intelligent based unemployment rate prediction in selected countries of Europe. International Journal of Finance & Economics, 2023, 28(1): 528-543.

[3] Marika Karanassou, Hector Sala. The rise and fall of Spanish unemployment: a chain reaction theory perspective. International Papers in Political Economy, 2009.

[4] Anil Kumar. Texas twist: why did state’s unemployment fall below nations. SouthwestEconomy, 2010, 3-7.

[5] Daniel Arkkelin. Using SPSS to understand research and data analysis. Valparaiso University, 2014.

[6] Katris C. Prediction of unemployment rates with time series and machine learning techniques. Comput. Econ., 2020, 55: 673–706.

[7] Chakraborty, T., Chakraborty, A.K., Biswas, M. et al. Unemployment rate forecasting: a hybrid approach. Comput Econ, 2021, 57: 183–201.

[8] Michał Gostkowski, Tomasz Rokicki. Forecasting the unemployment rate: application of selected prediction methods. European Research Studies Journal, 2021, 24(3): 985-1000.

[9] Sorin Burlacu, Amelia Diaconu, Petronela Evelina Bălu, Iulian Gole. The economic and social effects of unemployment in Romania. Revista de Management Comparat Internațional, 2021, 22(1): 21-27.

[10] Eriksson Stefan, Dan-Olof Rooth. Do employers use unemployment as a sorting criterion when hiring? Evidence from a field experiment. American Economic Review, 2014, 104 (3): 1014-1039.