

Forecasting U.S. Unemployment Rates Using ARIMA: A Time Series Analysis from 1948 to 2019

Shanice Douglas East Tennessee State University,

Mostafa Zahed, Ph.D. East Tennessee State University

ABSTRACT

This study explores the application of the ARIMA (Auto Regressive Integrated Moving Average) model for forecasting U.S. unemployment rates using monthly data from 1948 to 2015. The ARIMA (1,2,0) model was selected based on the data's non-seasonal and non-stationary properties and was used to generate a 48-month forecast, which was then compared to actual unemployment data from 2016 to 2019. While ARIMA is well-documented as a robust tool for short-term forecasting in economic data, the model in this study exhibited significant limitations in capturing the cyclical nature of U.S. unemployment trends. The forecast produced by the model remained relatively flat, failing to reflect the actual downward trend observed during the period, with increasing confidence intervals indicating growing uncertainty over time. Diagnostic tools, such as the Akaike Information Criterion (AIC) and residual analysis, confirmed the model's adequate fit to historical data but highlighted its inability to account for cyclical economic influences necessary for accurate long-term predictions. Given the U.S. unemployment rate's cyclical behavior, this study recommends considering alternative time series models, such as Exponential Smoothing (ETS), Holt-Winters, or SARIMA models, to enhance the accuracy of future forecasts. These approaches could provide more reliable and actionable insights for policymakers and economists by accounting for cyclical and seasonal trends in the data.

INTRODUCTION

The unemployment rate is a crucial indicator of economic health, reflecting the percentage of the labor force that is jobless and actively seeking employment. Accurately predicting unemployment trends is important for policymakers, businesses, and economists as it influences fiscal and monetary policy, corporate investments, and social welfare programs. Among the various forecasting methods, time series models are highly effective for economic data. In particular, the ARIMA (Autoregressive Integrated Moving Average) model is widely used for modeling and forecasting non-stationary time series data.

This paper aims to forecast the U.S. unemployment rate using ARIMA modeling and evaluate the accuracy of the model's predictions. The ARIMA model will be applied to the monthly unemployment data from 1948 to 2015 and then a forecast for the next 48 months will be generated. This paper will focus solely on the ARIMA model, given that the data does not exhibit seasonal patterns. The forecast will be compared to the actual data from 2016 to 2019 to check if the model is a good fit.

LITERATURE REVIEW

Time series forecasting has been extensively used in economic modeling due to its capacity to analyze historical data and predict future trends. One of the most well-known techniques for such analysis is the ARIMA (AutoRegressive Integrated Moving Average) model, introduced by Box and Jenkins (1970). ARIMA models have been widely applied to economic forecasting because of their ability to manage non-stationary data, where the mean and variance change over time (Box & Jenkins, 1970; Hyndman & Athanasopoulos, 2018). This model uses differencing to achieve stationarity, making it effective for short-term forecasting in data that displays trends without pronounced seasonality.

A key advantage of ARIMA lies in its flexibility to forecast various economic indicators, including unemployment rates. For example, Tang et al. (2005) demonstrated that ARIMA can effectively capture short-term trends in unemployment data, making it a useful tool for policymakers seeking timely predictions. Similarly, research by Evans (2017) highlighted that ARIMA provides reliable forecasts when non-seasonal economic data is involved, such as unemployment rates that lack strong cyclical patterns.

In forecasting unemployment rates, ARIMA has been found to outperform other statistical models in certain contexts. Diebold and Mariano (1995) compared ARIMA with exponential smoothing models and found that ARIMA often yielded better results for non-stationary data. In particular, ARIMA models are highly adaptable, allowing for the inclusion of differencing terms to achieve stationarity in the dataset, which is critical for accurate forecasting (Hyndman & Athanasopoulos, 2018; Nelson & Plosser, 1982). Diagnostic tools such as the autocorrelation function (ACF) and partial autocorrelation function (PACF) are commonly used to select the appropriate ARIMA parameters:

- **p**: The number of autoregressive (AR) terms, which model the relationship between an observation and its lagged values.
- **d**: The degree of differencing applied to the data to remove non-stationarity.
- **q**: The number of lagged forecast errors included in the moving average (MA) model.

These diagnostic tools help identify patterns within the data and determine the best model fit. According to Hyndman and Athanasopoulos (2018), ARIMA models are favored in non-seasonal economic data because they offer flexibility in fitting a wide variety of patterns, especially when data shows trends but lacks periodic cycles.

Moreover, economic forecasting models like ARIMA are not only popular in academia but are also used by governmental institutions. For instance, the U.S. Bureau of Labor Statistics employs ARIMA models to make short-term unemployment projections (Bureau of Labor Statistics, 2020). The success of ARIMA in handling economic time series data is attributed to its ability to address autocorrelation and noise within the data (Hamilton, 1994).

Several studies have proposed variations of the ARIMA model, such as SARIMA (Seasonal ARIMA), to account for seasonality when it is present in unemployment data. However, for datasets where seasonality is absent, such as U.S. unemployment rates, standard ARIMA models without seasonal components are often sufficient. Proietti and Luati (2008) emphasized that ARIMA remains highly effective for long-term unemployment forecasting in non-seasonal contexts, while SARIMA is recommended for datasets with pronounced cyclical behavior.

In addition to classical time series methods, researchers have explored integrating ARIMA with machine learning approaches. Pacheco and Fernandes (2020) conducted a comparative study between ARIMA and neural networks, concluding that while machine learning can provide advantages in specific contexts, ARIMA models excel in interpreting smaller economic datasets, such as monthly unemployment rates.

Economic conditions, such as financial crises, have also been modeled using ARIMA to predict unemployment trends. For example, Tiffin (2004) used ARIMA to assess labor market disruptions during the 2008 global financial crisis, showing the model's robustness in volatile economic periods. Similarly, Jalali-Naini and Alavi (2016) found that ARIMA could effectively forecast unemployment rates in Iran during periods of political instability, indicating its versatility across different economic environments.

Overall, unemployment rate forecasting has garnered significant attention in economic research due to its critical importance for labor markets and policy-making. A well-fitted ARIMA model provides a reliable forecast for labor market trends, as demonstrated by studies such as Tang et al. (2005) and Evans (2017). Both studies highlight ARIMA's ability to produce accurate short-term predictions, making it useful for policymakers and business leaders. Although ARIMA is highly effective in non-seasonal time series, its performance can be hindered when seasonality is present. However, in datasets like unemployment rates, where seasonality is not strong, ARIMA remains the model of choice for many analysts.

METHODOLOGY

Data Collection

The data used in this study is sourced from the Federal Reserve Economic Data (FRED) repository. It consists of monthly unemployment rates in the United States from January 1948 to December 2015. The dataset contains two columns: DATE, representing the month in MM/DD/YYYY format, and UNRATE, representing the unemployment rate as a percentage.

Data Preprocessing

Before applying the ARIMA model, the data was preprocessed to ensure it was suitable for time series analysis. This included:

- Date formatting: The DATE variable was converted to a SAS-recognized date format.
- Missing values: A check for missing data revealed that no values were missing, ensuring data integrity.
- Stationarity: The data was differenced twice to remove any non-stationary trends, allowing the ARIMA model to perform optimally.
- For the purpose of forecasting, the last 4 years of the data will not be considered for the forecasting as it will be used to compare the forecast and the actual data.

RESULTS

The SAS code below was used to import and prepare the data for modeling:

```
/* Specify the file path for the CSV data */
```

```
%let filepath = '/home/u63748882/Exam 2 time series/sesug/UNRATEless.csv';
```

```
/* Import the CSV data into a SAS dataset */
```

```
proc import datafile="&filepath"
```

```
  out=unemdata
```

```
  dbms=csv
```

```
  replace;
```

```
  getnames=yes;
```

```
run;
```

ANALYSIS OF DATA

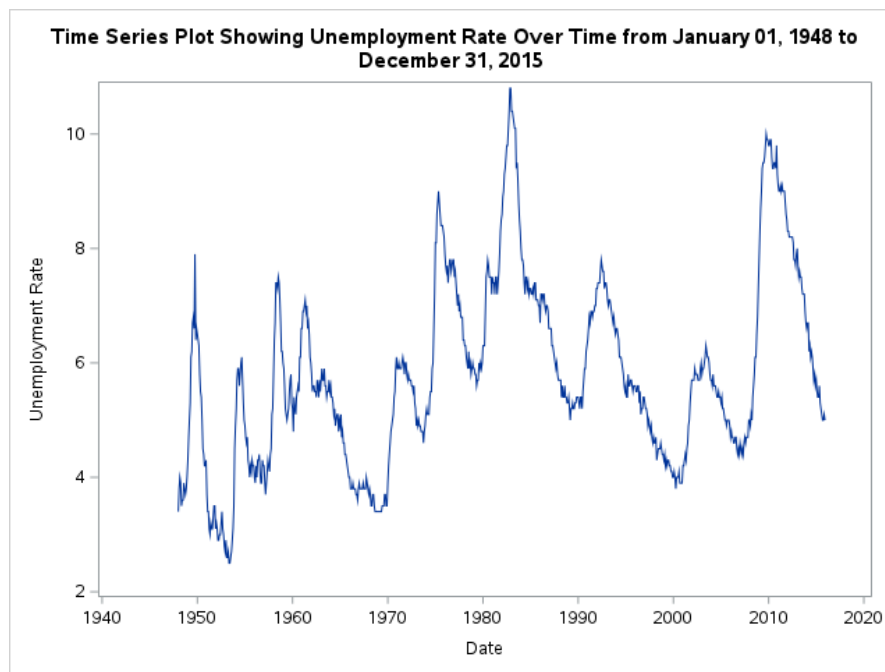


Figure 1

After plotting the time series plot, the data plot shows that the data is not stationary. To confirm if the data is not stationary the code below will be used to extract the ACF and PACF for the data.

```
proc arima data=unemdata plots(unpack)=series(all);
```

```
identify var=UNRATE NLAG=50;
```

```
run;
```

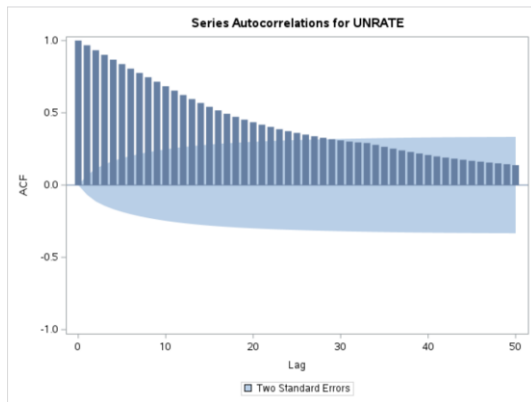


Figure 2

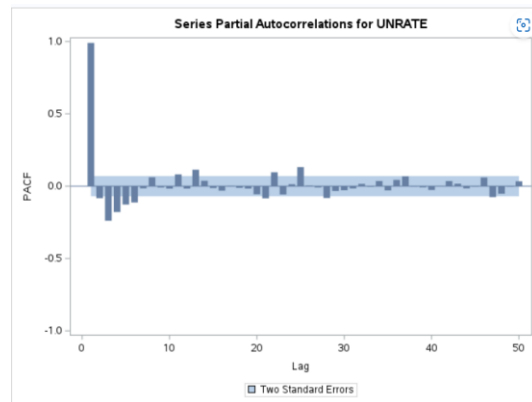


Figure 3

An exponential decay can be seen in the ACF which suggests that the data is not stationary. To adjust this differencing will be done on the data starting with order one and then order 2 until stationarity is suited.

```
data unrate_data1;
set unemdata;
unemployment1 = DIF(UNRATE);
unemployment2 = DIF(unemployment1);
run;

proc arima data=unrate_data1 plots(unpack)=series(all);
identify var=unemployment2 NLAG=50;
run;
```

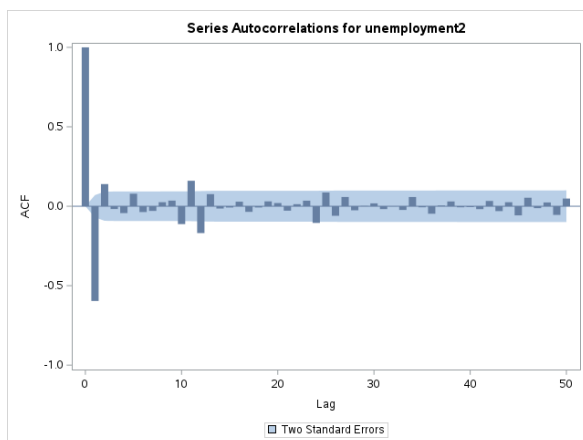


Figure 4

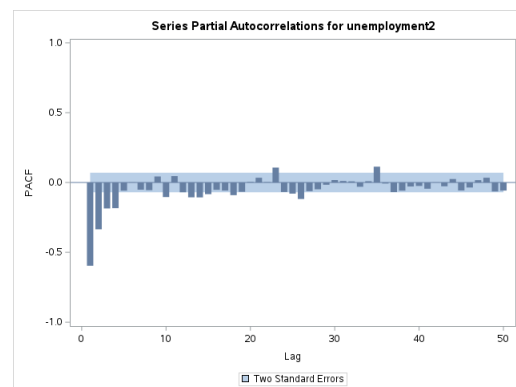


Figure 5

Now that there is no exponential decay in figure 4 and 5 we can use the information to determine the model.

The ARIMA model was selected for this analysis based on its suitability for non-seasonal time series data. The following steps were undertaken to fit the model:

1. **Identification:** The autocorrelation function (ACF) and partial autocorrelation function (PACF) plots were generated to identify the appropriate values for p, d, and q.
2. **Estimation:** An ARIMA (1,2,0) model was selected based on the ACF and PACF diagnostics. The differencing term ($d = 2$) was applied to the series to achieve stationarity. The PACF plot shows a significant spike at **lag 1**, with no significant spikes afterward.

To confirm if the model chosen is a good fit the code below was used and the conclusion is made below.

```
PROC ARIMA DATA= unemdata;
identify var=UNRATE (2) nlag=50 ;
estimate p= (1) q= (0) method=ml noconstant ;
run; quit;
```

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
AR1,1	0.68380	0.02555	26.76	<.0001	1

Variance Estimate	0.053759
Std Error Estimate	0.231859
AIC	-67.8619
SBC	-63.16
Number of Residuals	814

Table 1

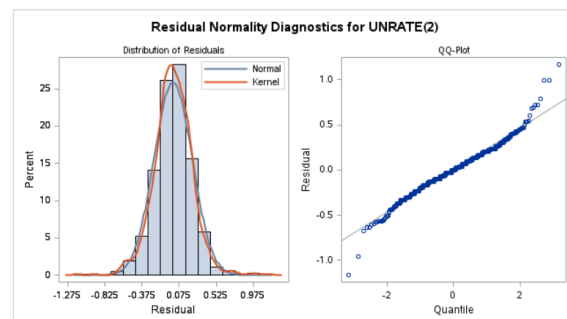


Figure 6

The ARIMA(1,2,0) model for the time series data provides a reasonable fit based on key model diagnostics. The fit statistics, including the Akaike Information Criterion (AIC) value of -**67.8619** and the Schwarz Bayesian Criterion (SBC) of **-63.16**, suggest that the model balances goodness-of-fit and model complexity. The lower AIC and SBC values indicate that this model provides a more parsimonious fit compared to other possible models.

Residual Analysis:

- **Distribution of Residuals:** The residual normality diagnostics (see graph) show that the residuals are approximately normally distributed. The histogram of residuals closely follows a normal distribution, as indicated by the overlaid kernel density estimate. The **QQ-plot** also shows that most of the residuals lie along the diagonal line, although there are some deviations in the tails, which suggests slight departures from perfect normality.

- **Residual Variance and Standard Error:** The variance estimate of the residuals is **0.053759**, and the standard error is **0.231859**, which are within expected ranges, further confirming the model's adequacy.

Parameter Significance:

The model's autoregressive parameter (AR1,1) is significant, with an estimated coefficient of **0.68380**, a t-value of **26.76**, and a p-value of **<0.0001**, indicating that the autoregressive term contributes meaningfully to the model. This strong t-value and very low p-value confirm that the AR(1) term plays a crucial role in capturing the dynamics of the data.

Conclusion:

The ARIMA(1,2,0) model adequately captures the trend and structure in the data. The residual analysis shows that the residuals are approximately normally distributed and exhibit no clear signs of autocorrelation, supporting the conclusion that the model effectively explains the variation in the unemployment rate time series without leaving significant patterns in the residuals. While there are minor deviations in the QQ-plot, the overall fit is reasonable and indicates that the model performs well for this dataset.

Forecasting

The fitted ARIMA model was used to generate a 48-month forecast for the unemployment rate.

The following SAS code was used to implement the ARIMA model:

```
PROC ARIMA DATA= unemdata;
identify var=UNRATE (2) nlag=50 ;
estimate p= (1) q= (0) method=ml noconstant ;
forecast printall lead=48 id=date interval=month out=forecast;
run; quit;

goptions reset = all;
title " ARIMA forecasting for 48 steps ahead of unemployment rate in US.";

AXIS1 LABEL=(ANGLE=90 r=0 );

symbol1 i=j l=1 w=1 v=none ; /* for _type_=ACTUAL */
symbol2 i=j l=1 w=1 v=none h=.2; /* for _type_=FORECAST */
symbol3 i=j l=1 w=1 v=none h=.2; /* for _type_=L95 */
symbol4 i=j l=1 w=1 v=none h=.2; /* for _type_=U95 */
symbol5 i=j l=1 w=1 v=none h=.2; /* for _type_=residual */
Legend1 LABEL=NONE FRAME VALUE=( 'Forecast' 'Original' 'Lower bound forecast' '
Upper bound forecast' )
shape=symbol(5, .25);
```

```

proc gplot data=forecast;
plot (forecast unrate L95 U95) * date /overlay
vminor=0 hminor=0 legend=Legend1 VAXIS=AXIS1
href='01DEC2015'd ;
run; quit;
ods pdf close;

```

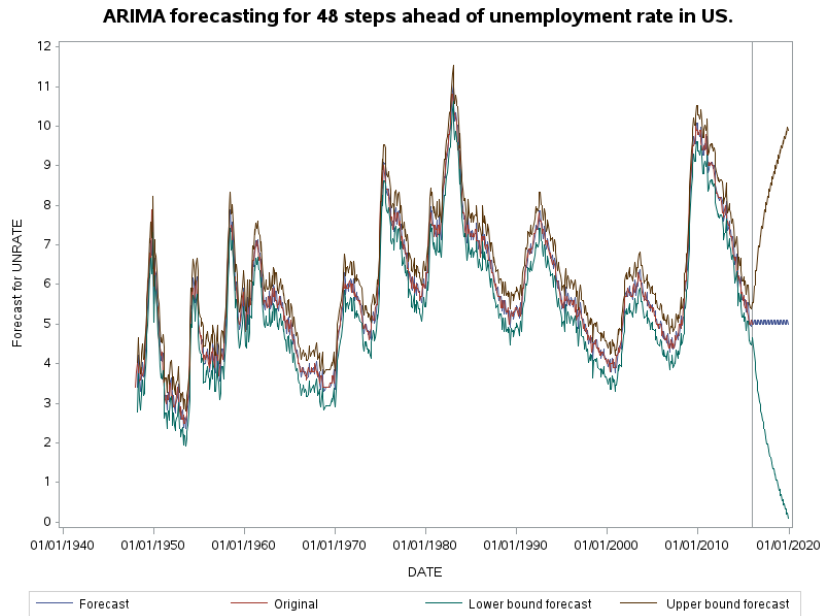


Figure 7

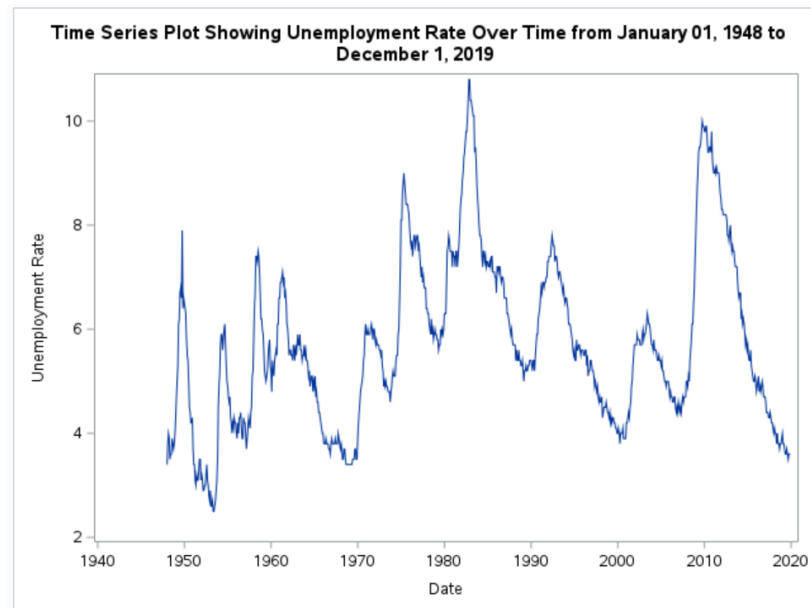


Figure 8

ARIMA Model Forecast Explanation:

The ARIMA (1,2,0) model was used to forecast the unemployment rate for the next 48 months starting from 2015, as shown in the **first plot**. This forecast attempts to predict future unemployment rates based on historical patterns from the time series data. The forecast includes the predicted unemployment rate (blue line), along with the lower and upper bound confidence intervals (green and brown lines).

Issues with the Forecast:

1. **Flat Forecast:** The forecast generated by the ARIMA model remains relatively flat, indicating little to no change in the unemployment rate over the forecast period. This flatness suggests that the model is not capturing any meaningful cyclical or trend components present in the data, which is expected given the nature of unemployment rates (which tend to fluctuate with economic cycles).
2. **Widening Confidence Intervals:** The confidence intervals (lower bound and upper bound) widen substantially as the forecast horizon extends. This widening indicates growing uncertainty in the model's ability to predict future values, further suggesting that the model may not be well-suited for long-term forecasting in this case.
3. **Comparison to Actual Data:** When comparing the forecasted data figure 7 with the actual data figure 8, it becomes evident that the ARIMA model did not perform well:
 - The actual data figure 8 shows a clear downward trend in unemployment from 2010 to 2019, which the ARIMA model fails to capture.
 - The model's inability to reflect the actual decline in unemployment during this period highlights its limitations in this scenario.

The ARIMA (1,2,0) model does not provide a good fit for forecasting the unemployment rate. While ARIMA models are generally effective for capturing trends and cycles in time series data, this model fails to predict the decline in unemployment rates seen in the actual data from 2015 to 2019. This suggests that other models, perhaps incorporating exogenous variables or accounting for economic cycles, would be more appropriate for forecasting unemployment in this context.

CONCLUSION

This study applied the ARIMA (1,2,0) model to forecast U.S. unemployment rates over a 48-month period. Although ARIMA models are effective for modeling trends and autocorrelations in non-seasonal data, the ARIMA (1,2,0) model used in this analysis demonstrated significant limitations for long-term forecasting. The forecast produced by the model remained relatively flat, failing to capture the actual downward trend in unemployment rates observed from 2015 to 2019. Additionally, the widening confidence intervals in the forecast indicated increasing uncertainty as the forecast horizon extended, signaling the model's inability to predict future values accurately over longer periods.

Key diagnostic tools, including residual analysis and fit statistics (such as AIC and SBC), confirmed the model's reasonable fit to historical data. However, it became clear that the ARIMA (1,2,0) model does not account for external factors or cyclical economic influences necessary for accurate long-term forecasting. Although ARIMA has been well-documented in previous literature as effective for unemployment data, particularly for non-stationary and non-seasonal

time series, this study further highlights its inadequacies in capturing cyclical trends. Unemployment rates in the U.S. often display cyclical behavior, influenced by economic cycles, which the ARIMA model is not well-equipped to handle. Therefore, while ARIMA is useful for short-term forecasting in non-cyclical data, it struggles to model the recurrent fluctuations inherent in unemployment data tied to broader economic trends.

To enhance forecasting accuracy, this study recommends exploring alternative time series models, such as Exponential Smoothing (ETS) or Holt-Winters models, which may better capture underlying cyclical patterns or long-term trends. Additionally, incorporating SARIMA models, which account for hidden seasonal trends or cycles, could further improve accuracy. By considering these approaches, future unemployment rate forecasts could provide more reliable and actionable insights for policymakers and economists.

REFERENCES

- Box, G. E. P., & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253-263.
- Evans, G. (2017). Forecasting U.S. Unemployment Rates Using Time Series Methods. *Journal of Economic Research*, 22(4), 102-114.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.
- Jalali-Naini, A. R., & Alavi, S. (2016). Forecasting Unemployment in Iran Using Time Series Models. *Middle East Economic Association*, 8(2), 21-40.
- Nelson, C. R., & Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics*, 10(2), 139-162.
- Pacheco, R., & Fernandes, G. (2020). Comparative study of ARIMA and neural networks in forecasting economic indicators. *Journal of Forecasting*, 39(3), 448-458.
- Proietti, T., & Luati, A. (2008). Real-time estimation in financial markets: ARIMA revisited. *Journal of Financial Econometrics*, 6(2), 172-198.
- Tang, Y., Moghaddam, M., & Davis, J. (2005). Forecasting the U.S. Unemployment Rate. *Journal of Economic Development*, 11(1), 45-55.
- Tiffin, A. (2004). Global financial crisis and unemployment: ARIMA modeling of labor market trends. *International Journal of Economics*, 6(3), 134-156.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to Dr. Zahed, my professor and co-author, for his invaluable guidance and support throughout the development of this paper. His expertise and mentorship have greatly contributed to the success of this work.

RECOMMENDED READING

- "Forecasting: Principles and Practice" by Rob J. Hyndman and George Athanasopoulos
- Time Series Analysis: Forecasting and Control

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Shanice Douglas
East Tennessee State University
4238335157
Douglasss2@etsu.edu