UNIVERSITY OF CONNECTICUT

OPIM 5671- DATA MINING AND BUSINESS INTELLIGENCE

Sudip Bhattacharjee

Time Series Forecasting Project Report.

“Unemployment Rate Prediction”

By:

Group 3

Nikhila

Raheem Abdul

Irfan Shaik

INDEX

1.Executive Summary:

This project focuses on the development and evaluation of Time Series models for forecasting unemployment rates using a dataset spanning from January 1, 1960, to July 1, 2023. The data set encompasses historical unemployment rates, federal interest rates and Job Openings. The primary objective was to create accurate unemployment rate predictions to aid in economic planning and policy formulation.

Upon initial data exploration, it was observed that there is no discernible trend or seasonality in the unemployment rate data. These findings challenge conventional time series forecasting methods and emphasize the importance of incorporating other relevant economic variables to improve prediction accuracy.

Several forecasting models were employed, including but not limited to autoregressive integrated moving average (ARIMA), exponential smoothing, and even including exogenous variable (ARIMAX). The models were evaluated based on metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) to identify the most effective approach.

Pre whitening: The application of pre whitening methods to the data, addressing autocorrelation, and rendering it more amenable to modeling.

Incorporating Job Openings: The inclusion of job openings as an additional predictor variable, recognizing its significance in the context of unemployment rate forecasting.

Time Series Analysis: Utilizing traditional time series models, including autoregressive integrated moving average (ARIMA), to develop forecasts for the unemployment rate.

Model Evaluation: Assessing the predictive performance of the models using appropriate metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

Notable findings from this project include:

The application of pre whitening techniques significantly improved the quality of the time series data, leading to more accurate forecasts.

The inclusion of job openings as a predictor variable demonstrated its relevance in unemployment rate forecasting, offering valuable insights into labor market dynamics.

In conclusion, this project contributes to the field of economic forecasting by demonstrating the effectiveness of machine learning techniques in predicting unemployment rates when traditional time series patterns are not apparent. It emphasizes the value of incorporating relevant economic variables, such as CPI inflation rates and federal interest rates, for more accurate predictions. The insights gained from this research have practical implications for policymakers, economists, and researchers interested in understanding and predicting labor market dynamics.

2.Introduction:

2.1 Project Overview:

This project focuses on forecasting unemployment rates using time series analysis. It tackles the challenge of predicting unemployment in the absence of clear trends or seasonality. Innovative techniques, including pre whitening, are employed to enhance data quality. Furthermore, the inclusion of job openings as a predictor variable enriches the predictive models. The project leverages a robust dataset spanning over six decades, from January 1, 1960, to July 1, 2023, encompassing unemployment rates, job openings. Its aim is to contribute insights that facilitate informed economic decisions and policy formulation.

2.2 Problem Statement:

The unemployment rate is a vital metric utilized by governments, economists, and researchers to assess labor market and economic well-being. To clarify, it is determined by expressing the number of unemployed persons as a percentage of the total number of persons in the labor force.

In this analysis, we aim to construct a model for the United States' unemployment rate. Our analysis will encompass an exploration of historical economic data, examining trends, seasonality, significant events, and various contributing factors. Ultimately, we will employ this model to forecast the near-term unemployment rate, offering valuable insights into the economic landscape.

2.3 Data Sources:

Our primary data source is the U.S FRED Economic Research Institute. This dataset encompasses monthly data on various labor market indicators. We'll focus on the unemployment rate as the primary target variable and may incorporate other relevant economic indicators like Job openings, Federal interest rate, NASDAQ and more.

Data Source: https://data.oecd.org/unemp/unemployment-rate.htm

https://fred.stlouisfed.org/series/UNRATENSA

By leveraging this data, we aim to gain insights into unemployment rate patterns, identify any underlying trends or seasonality, and provide reliable forecasts to take informed decisions for policy makers.

2.4 Data Description:

The data set records the monthly unemployment rate from 1st January 1960 to 1st July 2023. The data contains 63 years of data of Country USA.

The data set typically consists of 763 Rows and 4 columns.

The Dataset contains the below labels:

● Time Series (mm/dd/yyyy) - Date of 1st of every month.

● UNRATE (Dependent variable) – The value for every month

● Job Opening (exogenous value) - Measure of unfilled jobs on the last business day of the month.

● Federal Interest rate (exogenous value) – Overnight interest rate for interbank trading of federal funds

3.Data Preprocessing:

The dataset used for this analysis was sourced from the Fred Solutions. Observations are recorded on the 1st day of every month for over 63 years in USA.

3.1 Data Cleaning and transformation:

Our dataset exhibits missing values in the "job openings" column, specifically during the years 1960 to 2000. To address this issue, we need to perform imputation techniques to handle these gaps in the data. Imputation methods such as interpolation or regression analysis can help us estimate and fill in these missing values, ensuring the completeness and integrity of our time series dataset for further analysis and insights.

3.2 Handling Missing Values:

In our dataset there are missing values in the job openings from 1960 to 2000 year. So, to handle the missing values we have used the median value of accumulated time series to imputation in the Data Preparation step.

4. Time series Exploration:

This was our first approach to the model to look for patterns, trends, or any interesting features about the data.

4.1 Trend Analysis:

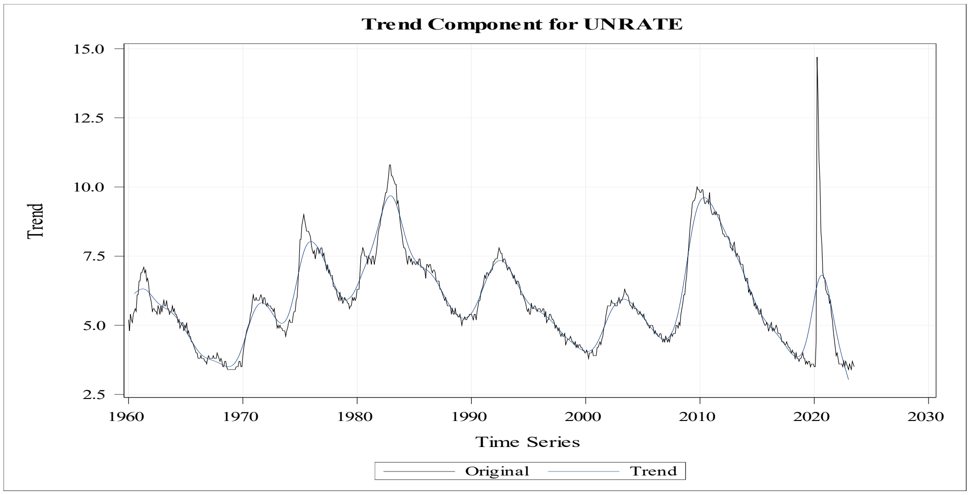


Fig:4.1 Trend component for Unemployment Rate

By looking into the graph, we found that there is no specific trend in the data. In 2020 because of the covid unemployment rate has been increased and in the year 1983 USA was faced the big recession due to that effect unemployment rate has been increased.

4.2 Seasonality and Patterns:

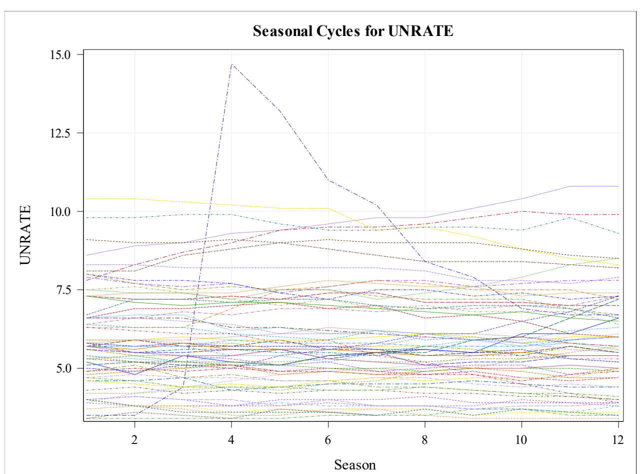


Fig 4.2 Seasonal Component for Unemployment rate

The seasonal component plot clearly tells that there is no specific seasonality in the data. We can clearly observe that there is no specific cycles in the data.

4.3 Interpretation:

This time series dataset consists of the official U.S. unemployment statistics, generated by the U.S. Bureau of Labor Statistics, and made available on the first Friday of each month. This dataset covers the period from 1960 to July 2023, providing insights into the fluctuations in the unemployment rate across various historical events.

The pronounced spikes in our time series data are primarily a result of recessions triggered by unexpected events occurring in those respective years. As a consequence, this result in our time series lacks discernible trends and seasonality patterns. it showcases how the unemployment rate has oscillated during a variety of historical events.

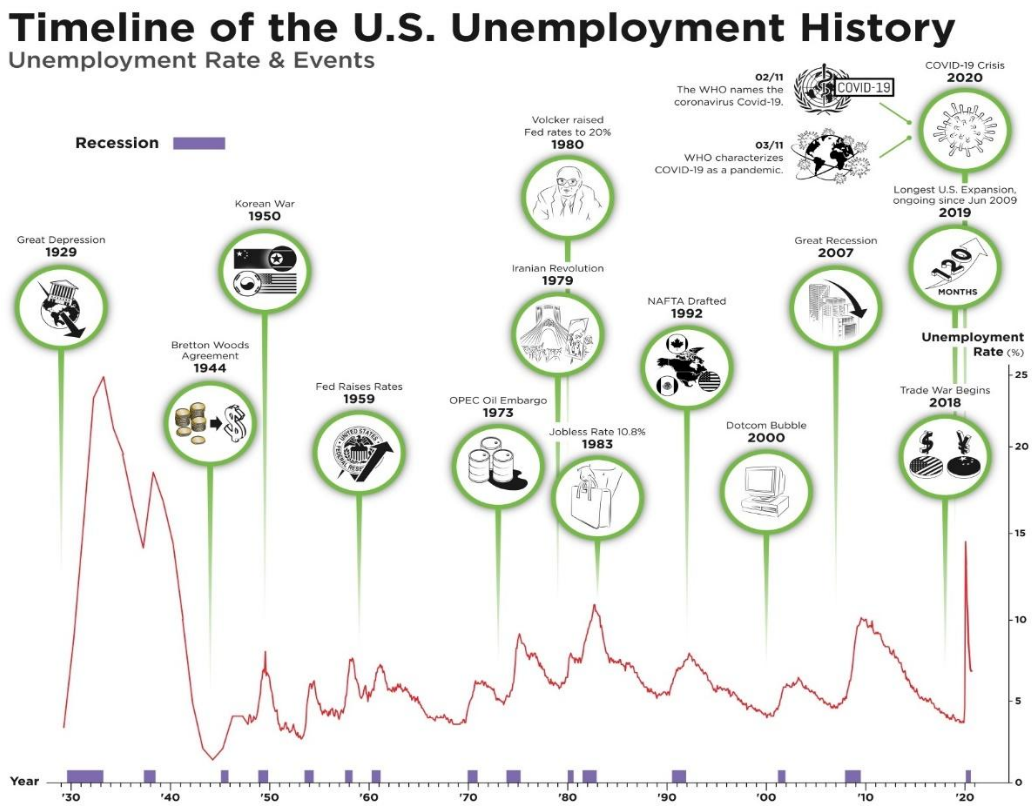


Fig 4.3 History of Un employment rate in the US

4.4 Test for Stationarity:

Augmented Dickey-Fuller Unit Root Tests							
Type	**Lags**	**Rho**	**Pr < Rho**	**Tau**	**Pr < Tau**	**F**	**Pr > F**
Zero Mean	0	-2.0750	0.3221	-1.07	0.2575		
	1	-2.1415	0.3146	-1.07	0.2578		
	2	-1.9681	0.3345	-1.05	0.2648		
Single Mean	0	-24.1894	0.0042	-3.43	0.0105	5.91	0.0130
	1	-26.7903	0.0024	-3.59	0.0063	6.47	0.0010
	2	-23.7842	0.0047	-3.34	0.0140	5.58	0.0204
Trend	0	-24.1656	0.0284	-3.43	0.0486	5.98	0.0647
	1	-26.7501	0.0161	-3.59	0.0316	6.57	0.0413
	2	-23.7424	0.0311	-3.33	0.0623	5.65	0.0831

Fig 4.4 Dickey-Fuller Test Results

By the augmented Dickey-Fuller Unit test we can observe that the data is slightly stationary. The Pr>F is significantly different from 0. Tau test p values are less than 0.05. so the null hypothesis can be rejected and hence the data is stationary.

4.5 White noise Probability Test:

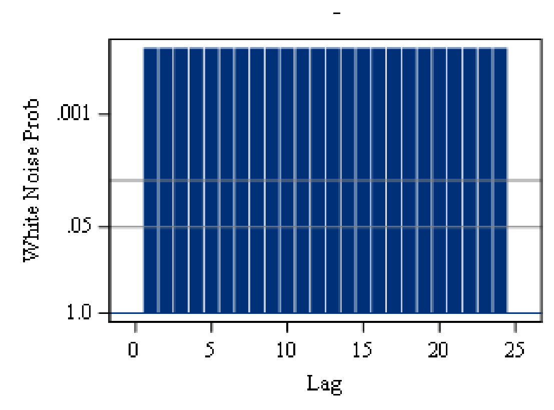


Fig 4.5 White Noise test in the entire data

Analyzing the White Noise probability graph, it becomes evident that there is a concentrated signal and an absence of White Noise. This suggests the potential for constructing models that can distribute errors resembling White Noise. Furthermore, it implies the possibility of extracting additional data from the given information.

4.6 Correlation graph:

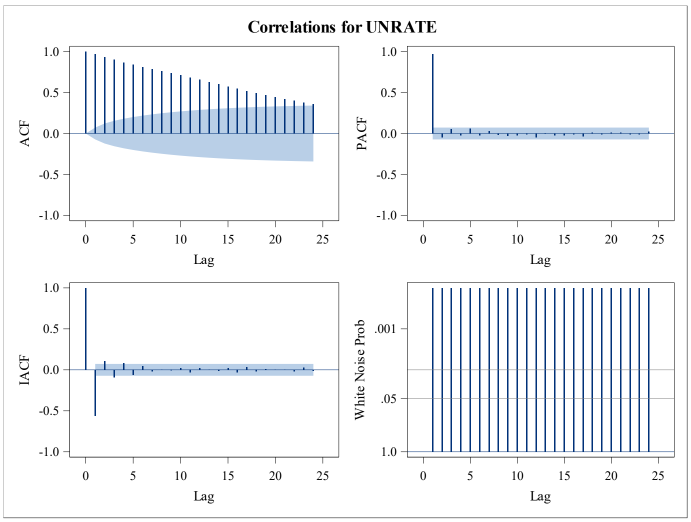


Fig 4.6 Correlations Graph

From the Correlation graph, we see from the ACF plot, the ACF decays very slowly, and the IACF have the spike at the lag2. So, this could be an Auto regression Model. So, by these observations we are going with the ARMA (2,0,0).

5. Model Development:

5.1 Model 1: - ARMA (2,0,0)

By examining the correlations plot with non-significant p, d, and q values, we can derive the appropriate orders for autoregressive (AR) and moving average (MA) components from the AutoCorrelation Function (ACF), Partial AutoCorrelation Function (PACF), and Integrated AutoCorrelation Function (IACF). This assessment is conducted under the assumption of stationarity.

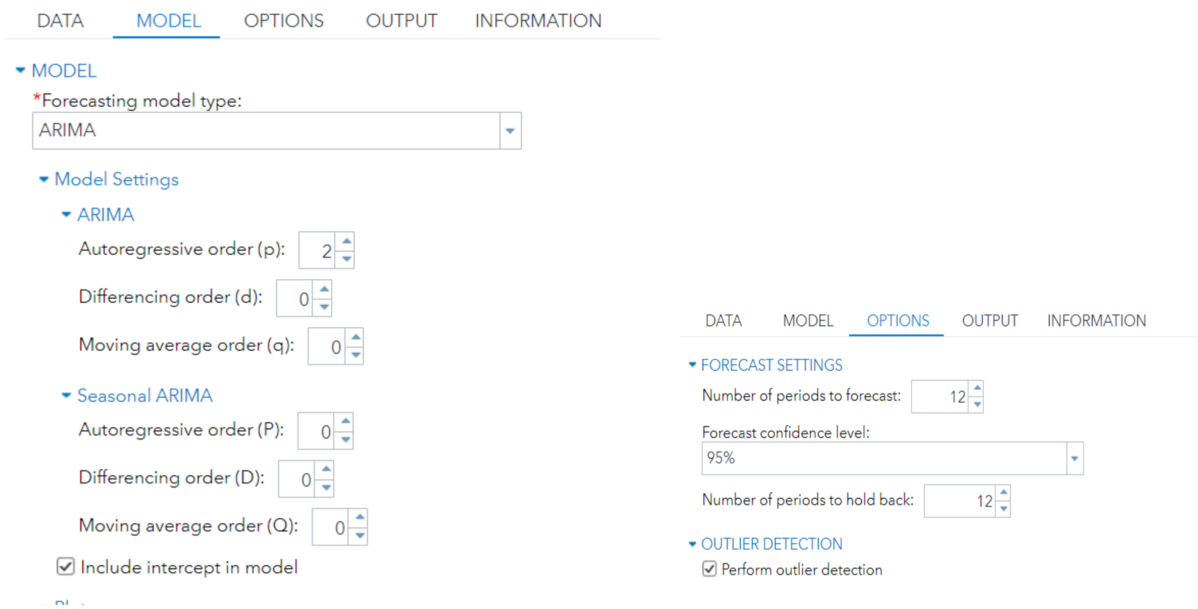


Fig 5.1 Model Building

The ARMA with (2,0,0) is performed and below given are the results of how the correlations and the plots are defined for it.

Model Results:

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	5.82807	0.43756	13.32	<.0001	0
AR1,1	1.01336	0.03617	28.02	<.0001	1
AR1,2	-0.04775	0.03622	-1.32	0.1874	2

Constant Estimate	0.200411
Variance Estimate	0.185114
Std Error Estimate	0.430249
AIC	884.0231
SBC	897.9349
Number of Residuals	763

Fig 5.2 Maximum Likelihood Estimates

Here in this observation, we can understand that only AR1,1 is significant and AR1,2 is insignificant.

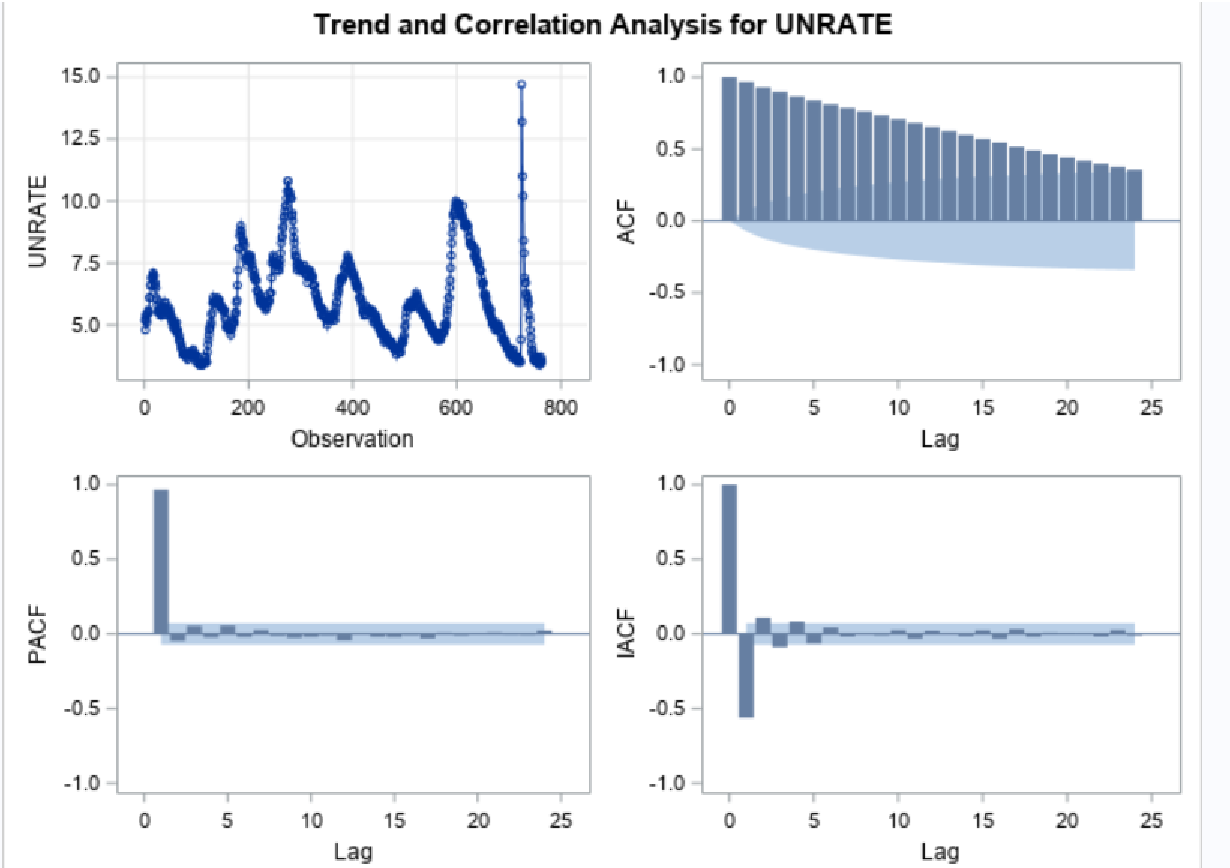


Fig 5.3 Correlation Graph

The ACF graph is clearly decreasing and the PACF and the IACF values are showing a good correlation. Furthermore, we can refer to the residual correlation to check for white noise signal whether it is captured or not to fit the model at the best.

Correlations of Parameter Estimates			
Parameter	MU	AR1,1	AR1,2
MU	1.000	0.000	-0.013
AR1,1	0.000	1.000	-0.967
AR1,2	-0.013	-0.967	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	4.90	4	0.2977	0.004	-0.051	0.021	-0.049	0.025	-0.017
12	8.39	10	0.5911	0.013	0.038	0.023	0.006	0.048	-0.006
18	9.59	16	0.8870	0.014	0.019	-0.001	0.028	-0.013	0.005
24	10.43	22	0.9820	-0.010	-0.015	0.009	0.009	-0.019	-0.014
30	10.63	28	0.9988	0.008	-0.000	-0.002	0.000	-0.003	0.014
36	11.88	34	0.9998	-0.015	0.004	-0.019	0.013	-0.007	-0.027
42	12.35	40	1.0000	0.006	-0.005	-0.016	0.001	-0.015	-0.008
48	13.53	46	1.0000	0.016	0.006	-0.012	0.015	-0.025	0.012

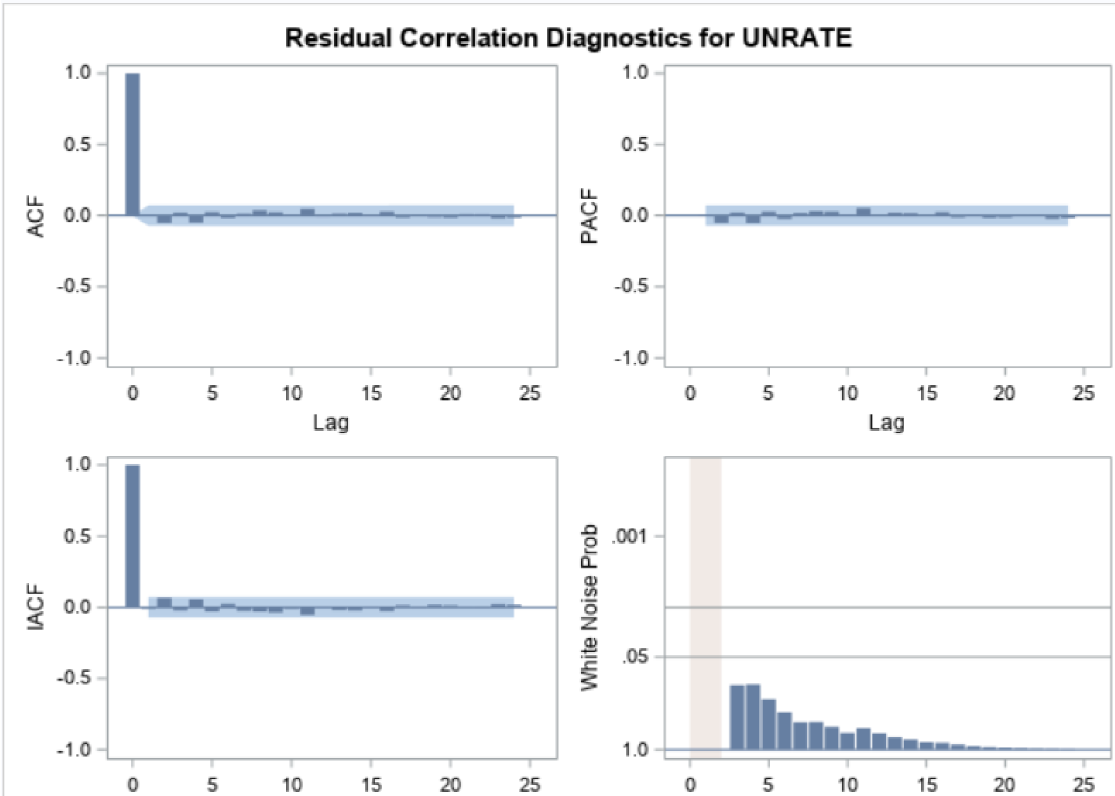


Fig 5.4 White Noise test Observation

Based on the white noise probability test, it appears that there is no signal in the error.

In the white Noise test, there is no white noise in the Residuals which defines the randomness of the data.

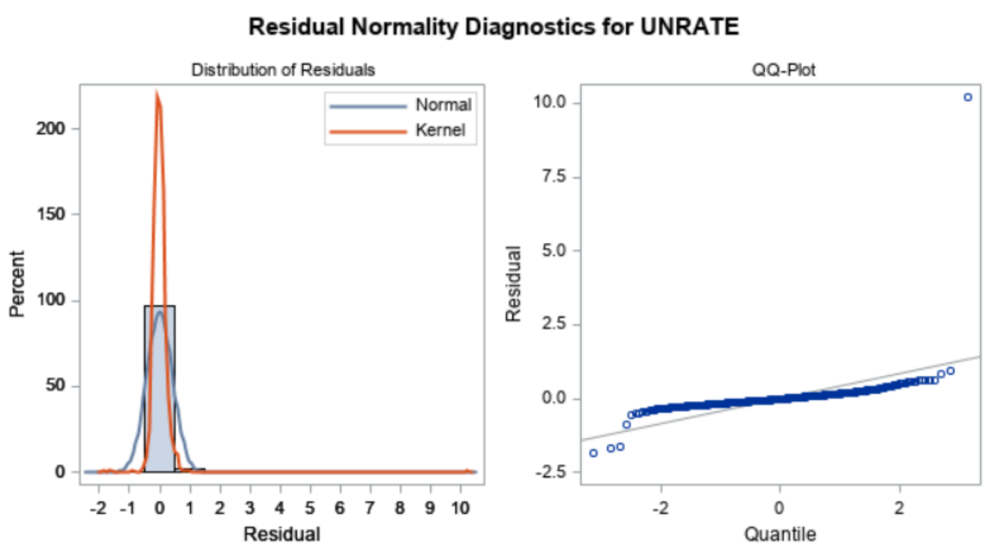


Fig 5.5 Residual Q-Q Plot

Departures from a straight line in the QQ plot are deviating from the diagonal line which mean the Skewness in the data.

Forecasting:

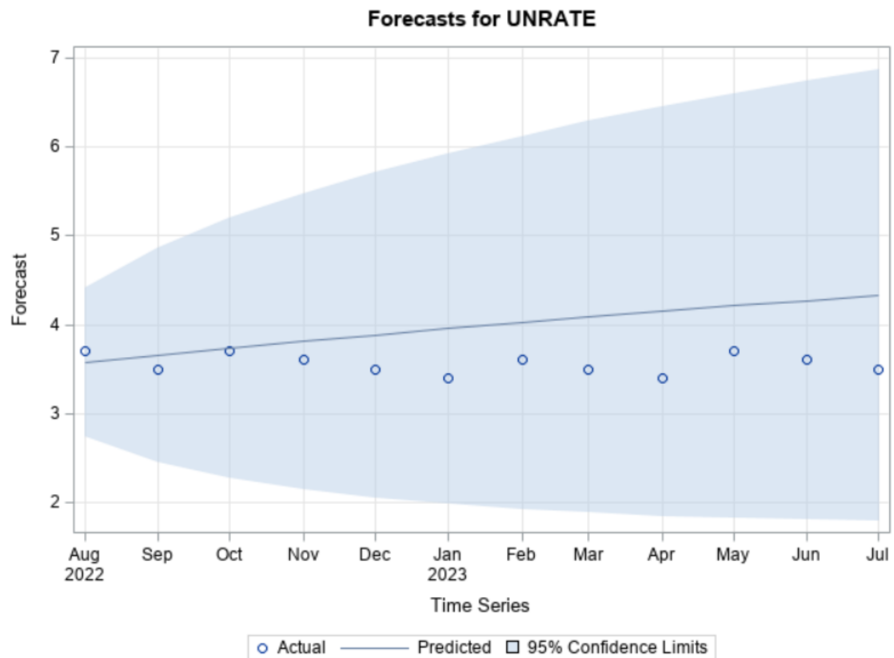
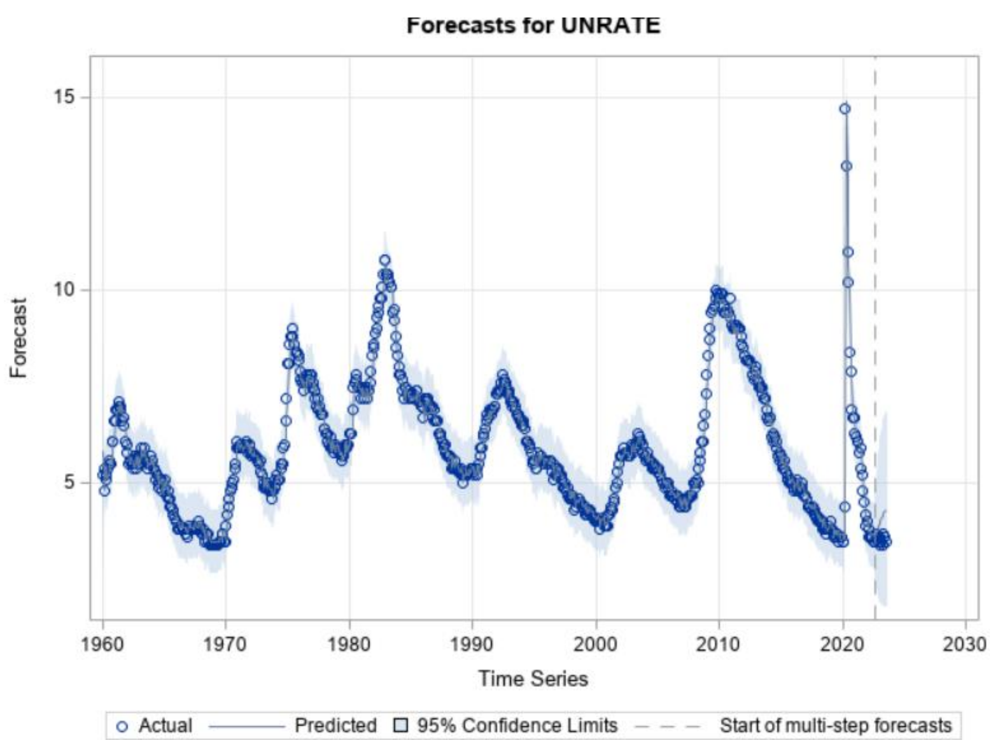


Fig:5.6 Forecasting Results



Fit Statistics:

Total rows: 11 Total columns: 3

	TYPE	_STAT_	_VALUE_
1	ML	AIC	884.02312501
2	ML	SBC	897.9348991

Fig 5.7 Fit Statistics

Using the Macros to derive the accuracy for the Model.

```
/* STSM03s04c.sas */


%let nhold=12;
%include "C:\Study\Sem-2(Fall)\Data Mining\forecasting\Forecasting-datasets-OPIM5671\Macros2.sas" / source2;
%accuracy_prep(indsn=stsm.'DM FC PRJ'n, series=UNRATE, timeid='Time Series'n,
    numholdback=&nhold);

```

```
proc arima data=WORK._TEMP plots=none;
    identify var=_y_fit;
    estimate p=(1 2) method=ML;
    forecast lead=&nhold id='Time Series'n interval=month out=WORK.OUTARM2 nooutall;
    run;
```

```
%accuracy(indsn=WORK.OUTARM2, timeid='Time Series'n, series=UNRATE,
    numholdback=&nhold);
```

Total rows: 1 Total columns: 7 Rows 1-1

	Model	Series	Holdback Periods	MAPE	MAE	MSE	RMSE
1	WORK.OUTARM2	UNRATE	12	12.95%	0.4561573714	0.2744836334	0.5239118565

Fig5.8 Accuracy Macro

Achieving 87.05% accuracy using ARMA (2,0,0) model.

5.2 Model 2: - ARIMA (1,1,0)

In the second model we tried with the differencing order 1 to make the data stationary but the results are like

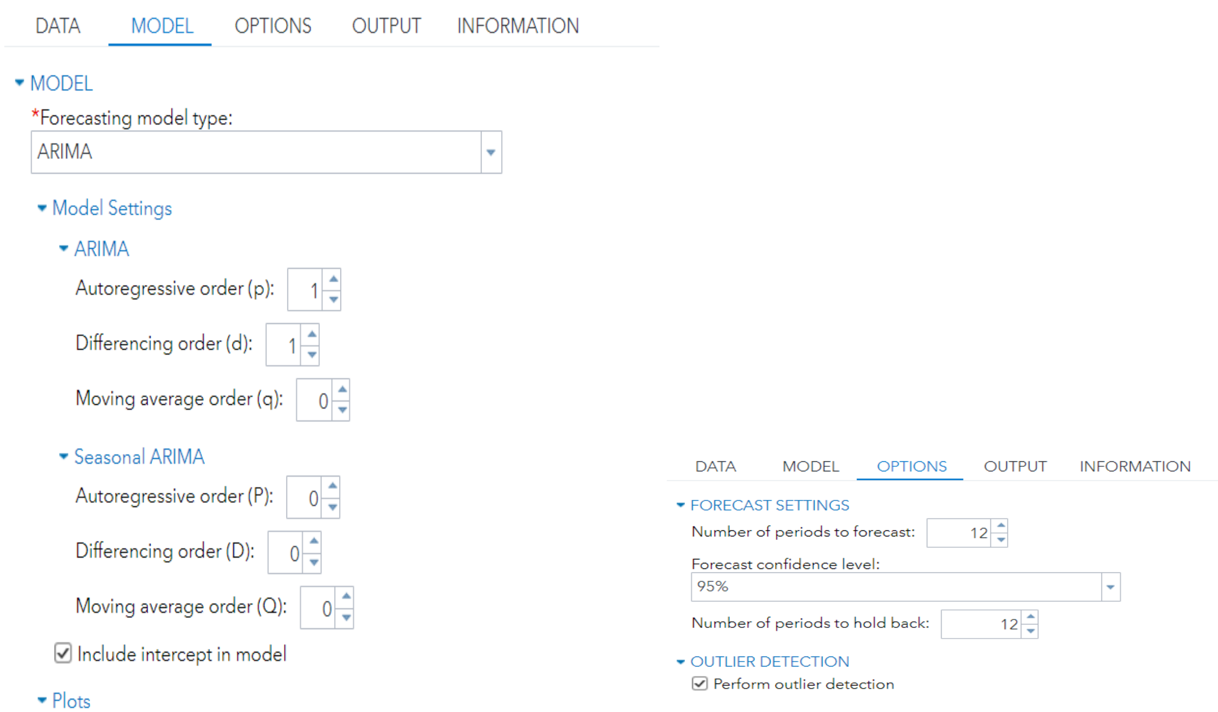


Fig: 5.9 Model Fitting

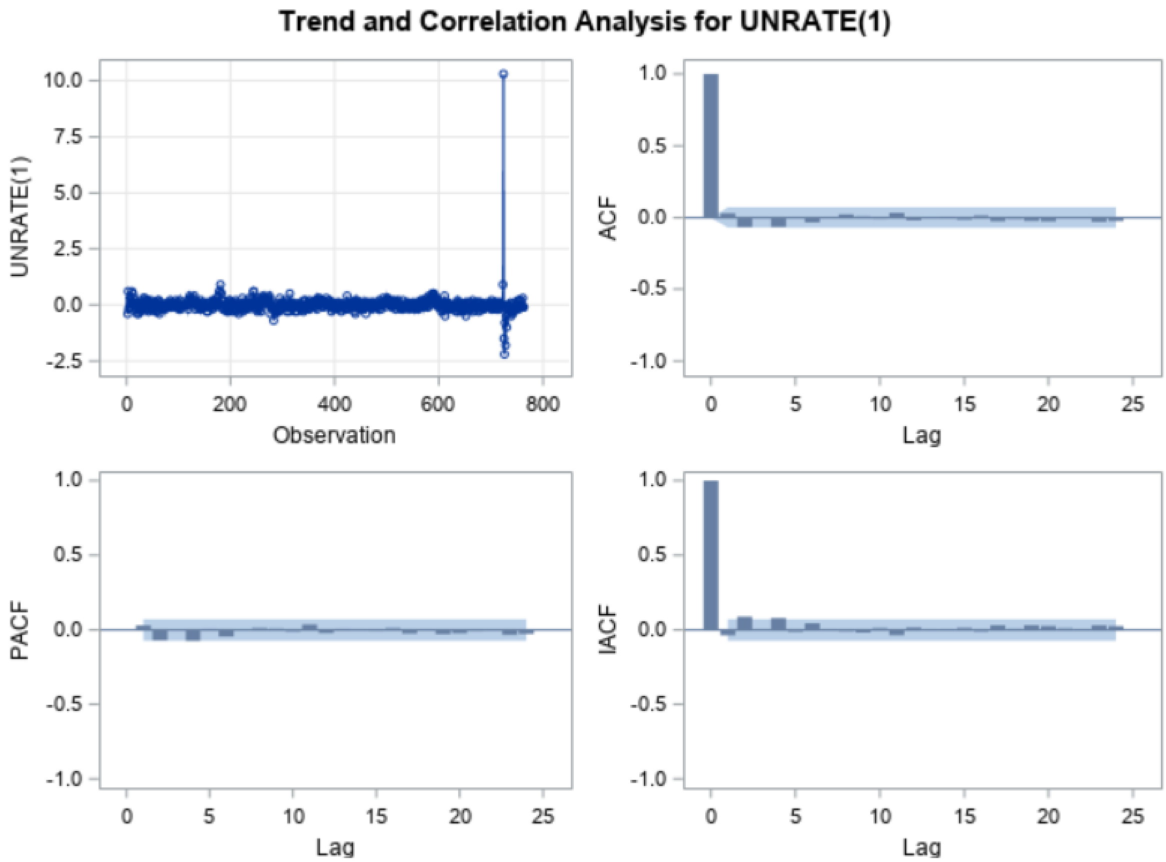
Model Results:

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.0022514	0.01621	-0.14	0.8895	0
AR1,1	0.03050	0.03626	0.84	0.4002	1

Constant Estimate	-0.00218
Variance Estimate	0.188201
Std Error Estimate	0.433821
AIC	891.7345
SBC	901.0064
Number of Residuals	762

Fig 5.10 Maximum Likelihood Estimates

Here in this observation, we can understand that mean (MU) & AR1,1 is insignificant. When the differencing order is 1.



Correlations of Parameter Estimates		
Parameter	MU	AR1,1
MU	1.000	-0.000
AR1,1	-0.000	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	8.09	5	0.1514	0.002	-0.069	0.001	-0.067	0.007	-0.034
12	9.86	11	0.5433	-0.003	0.023	0.010	-0.008	0.035	-0.019
18	10.75	17	0.8695	0.001	0.007	-0.013	0.016	-0.025	-0.006
24	12.57	23	0.9608	-0.021	-0.024	0.000	-0.000	-0.028	-0.022
30	12.78	29	0.9960	0.001	-0.007	-0.008	-0.006	-0.008	0.009
36	14.32	35	0.9992	-0.019	-0.001	-0.023	0.010	-0.010	-0.029
42	14.97	41	0.9999	0.005	-0.005	-0.015	-0.003	-0.020	-0.012
48	16.15	47	1.0000	0.012	0.002	-0.016	0.011	-0.029	0.009

Fig 5.11 Model Fitting Statistics

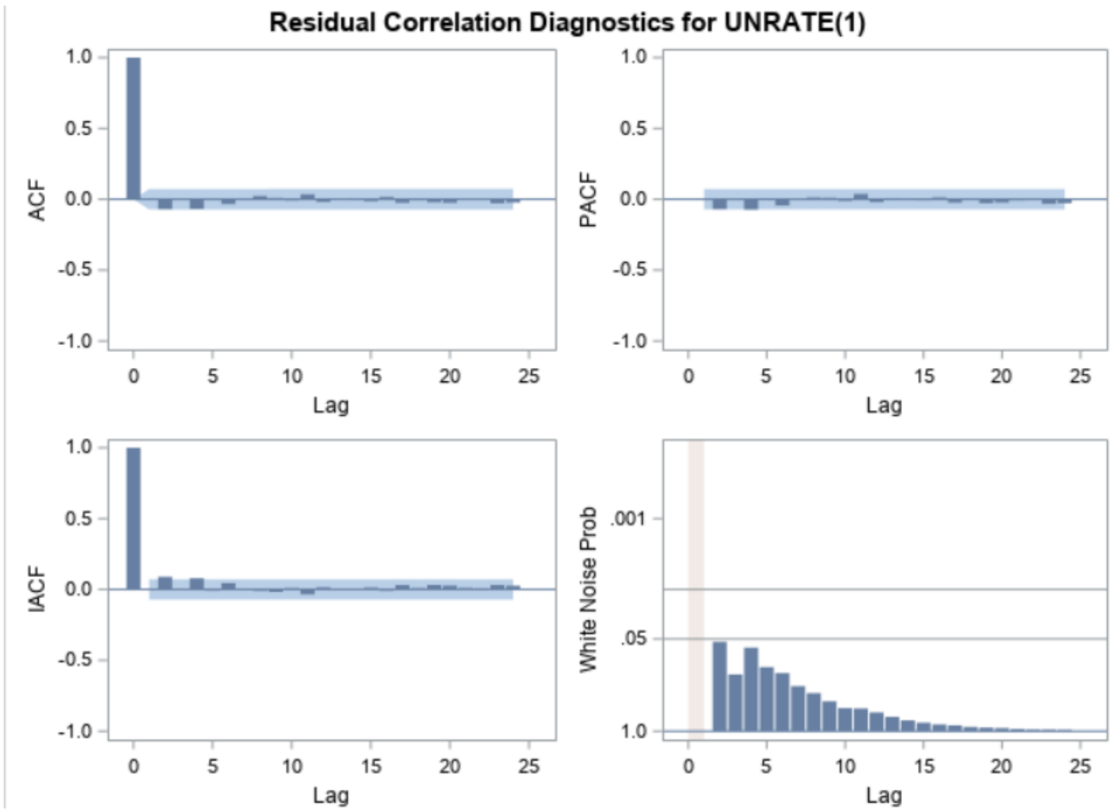


Fig 5.12 White Noise test Observation

Based on the white noise probability test, it appears that there is no signal in the error.

In the white Noise test, there is no white noise in the Residuals which defines the randomness of the data.

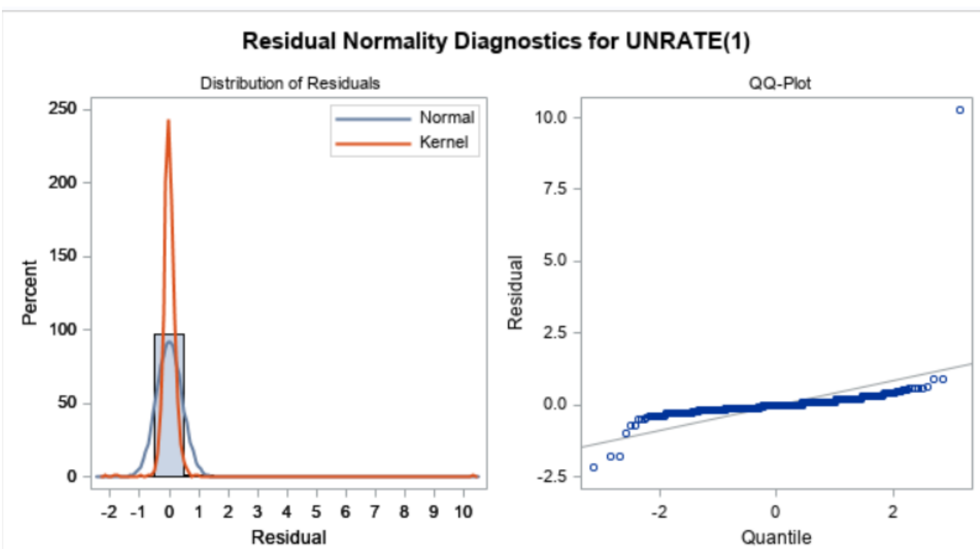


Fig 5.13 Residual Q-Q Plot

Departures from a straight line in the QQ plot are deviating from the diagonal line which mean the Skewness in the data.

Forecasting:

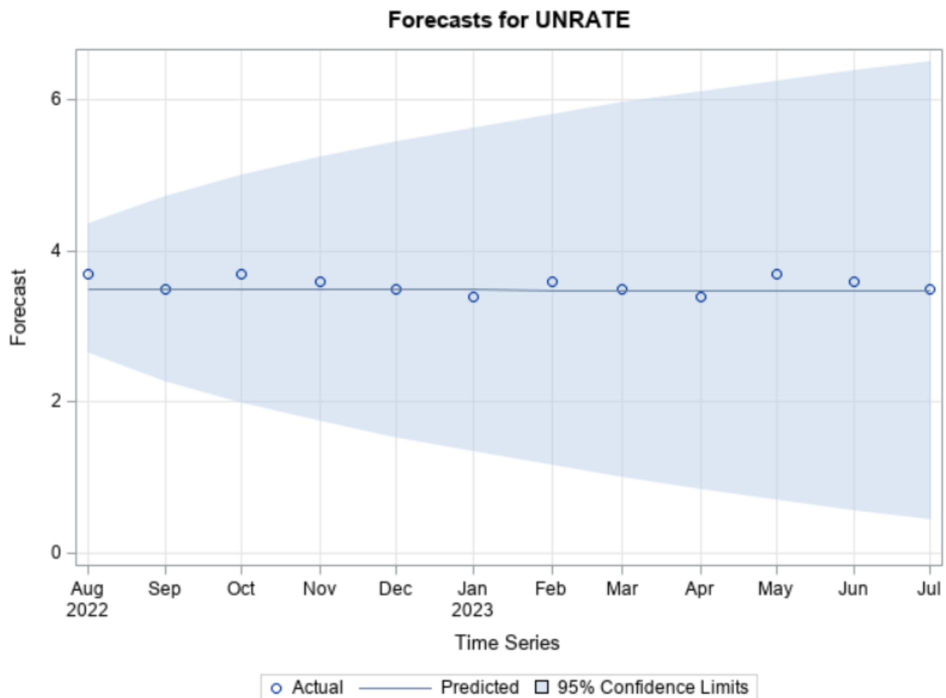
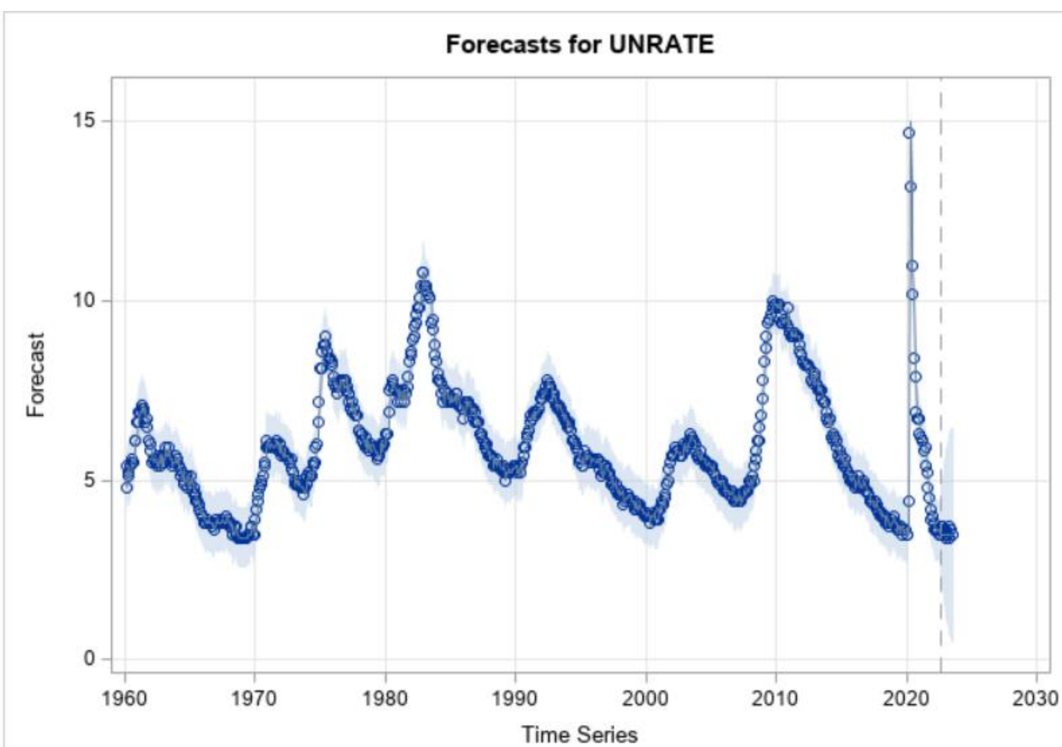


Fig 5.14 Forecasting Results



Fit Statistics:

Total rows: 11 Total columns: 3

	TYPE	_STAT_	_VALUE_
1	ML	AIC	891.7344814
2	ML	SBC	901.00637451

Fig 5.15 Fit Statistics

Using the Macros to derive the accuracy for the Model.

```
/* STSM03s04c.sas */


%let nhold=12;
%include "C:\Study\Sem-2(Fall)\Data Mining\forecasting\Forecasting-datasets-OPIM5671\Macros2.sas" / source2;
%accuracy_prep(indsn=stsm.'DM FC PRJ'n, series=UNRATE, timeid='Time Series'n,
    numholdback=&nhold);

```

```
proc arima data=WORK._TEMP plots=none;
    identify var=_y_fit(1);
    estimate p=(1) method=ML;
    forecast lead=&nhold id='Time Series'n interval=month out=WORK.OUTARM11 nooutall;
    run;
```

```

%accuracy(indsn=WORK.outARM11, timeid='Time Series'n, series=UNRATE,
     numholdback=&nhold);

```

Series	Model	Holdback Periods	MAPE	MAE	MSE	RMSE
UNRATE	WORK.OUTARM11	12	2.85%	0.10295	0.016262	0.12752

Fig 5.16 Accuracy Macro

Achieving 97.15% accuracy using ARIMA (1,1,0) model.

5.3 Model 3: - ARMA (1,0,0):

In the model 3 we built the model without adding differencing order 1 to check whether it is giving the accurate results.

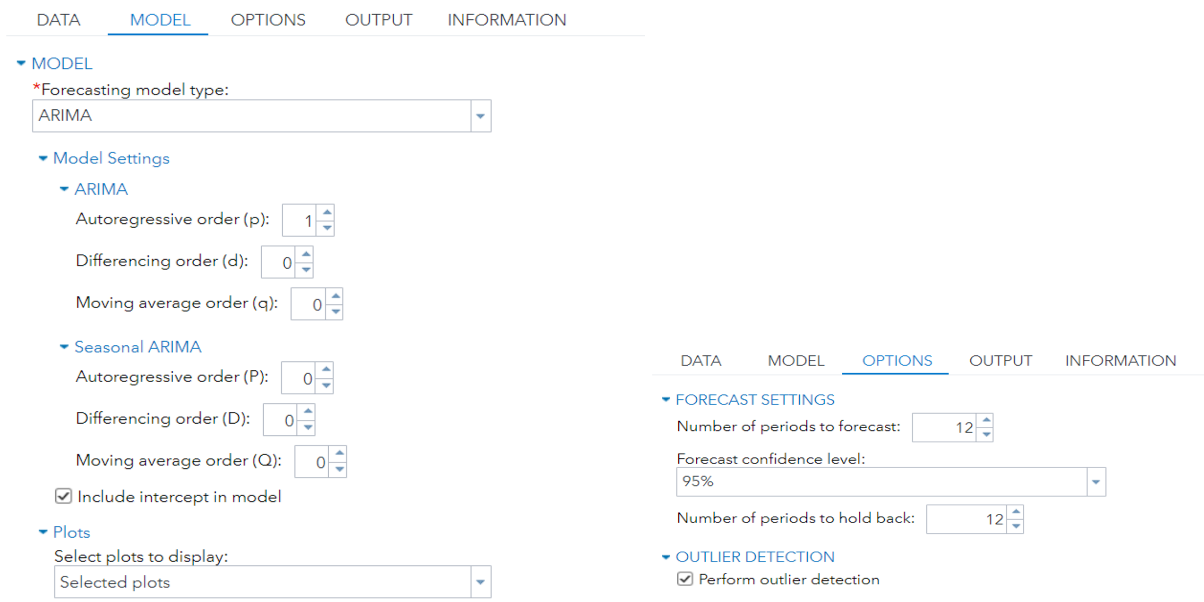


Fig 5.17 Model Fitting

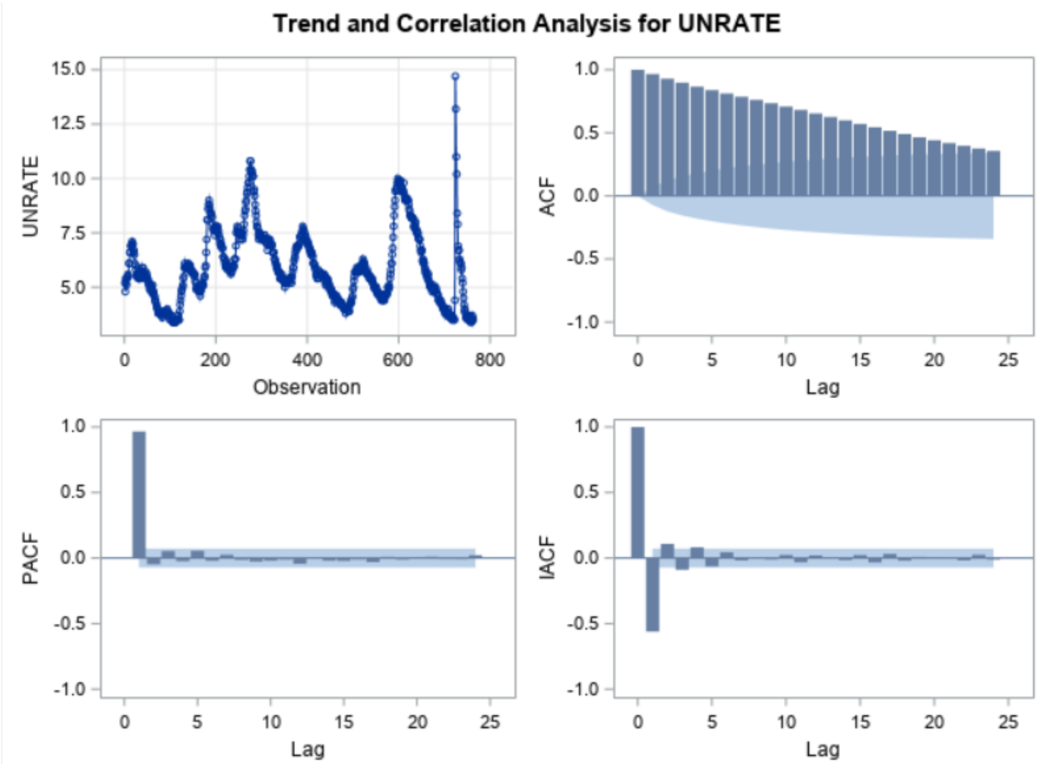
Model Results:

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	5.81752	0.45739	12.72	<.0001	0
AR1,1	0.96720	0.0092025	105.10	<.0001	1

Constant Estimate	0.190839
Variance Estimate	0.185293
Std Error Estimate	0.430457
AIC	883.7613
SBC	893.0358
Number of Residuals	763

Fig 5.18 Maximum Likelihood Estimates

Here in this observation, we can understand that mean (MU) & AR1,1 is significant.



Correlations of Parameter Estimates		
Parameter	MU	AR1,1
MU	1.000	-0.053
AR1,1	-0.053	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	6.35	5	0.2739	0.048	-0.051	0.013	-0.050	0.018	-0.018
12	9.56	11	0.5704	0.011	0.037	0.023	0.007	0.046	-0.005
18	10.57	17	0.8781	0.012	0.018	-0.001	0.025	-0.014	0.002
24	11.54	23	0.9770	-0.012	-0.016	0.007	0.007	-0.021	-0.016
30	11.70	29	0.9982	0.006	-0.001	-0.003	-0.001	-0.003	0.012
36	12.95	35	0.9998	-0.015	0.001	-0.019	0.011	-0.008	-0.028
42	13.46	41	1.0000	0.004	-0.006	-0.016	-0.001	-0.016	-0.008
48	14.50	47	1.0000	0.015	0.005	-0.012	0.013	-0.025	0.010

Fig 5.19 Model Fitting Statistics

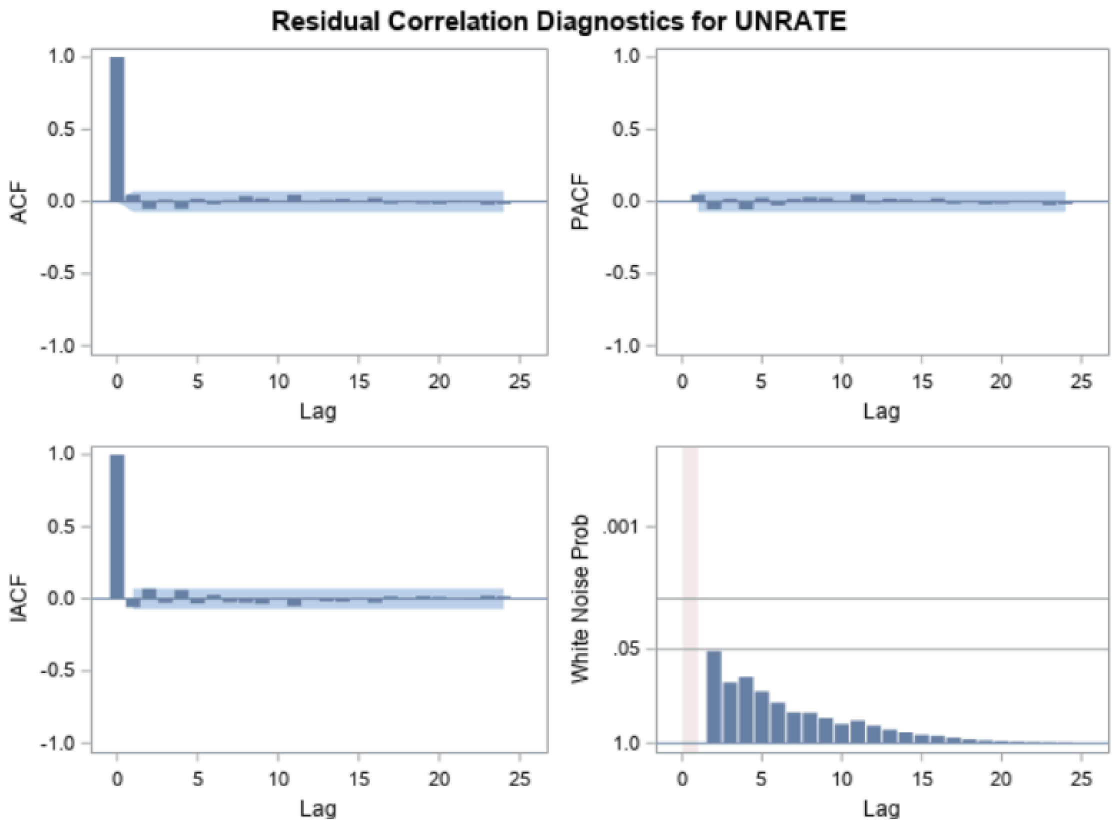


Fig 5.20 White Noise test Observation

Based on the white noise probability test, it appears that there is no signal in the error.

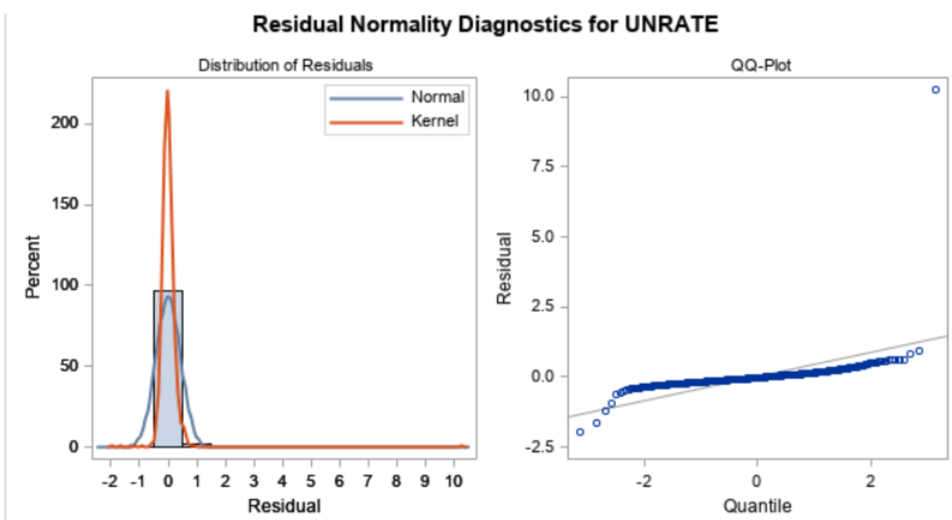
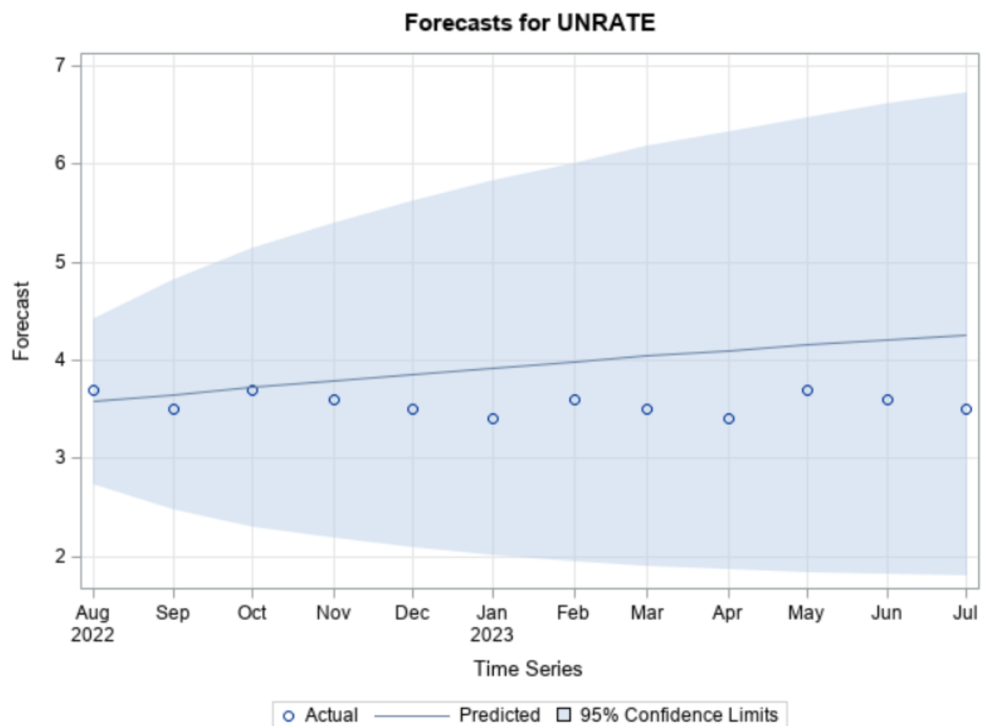


Fig 5.21 Residual Q-Q Plot

Forecasting:



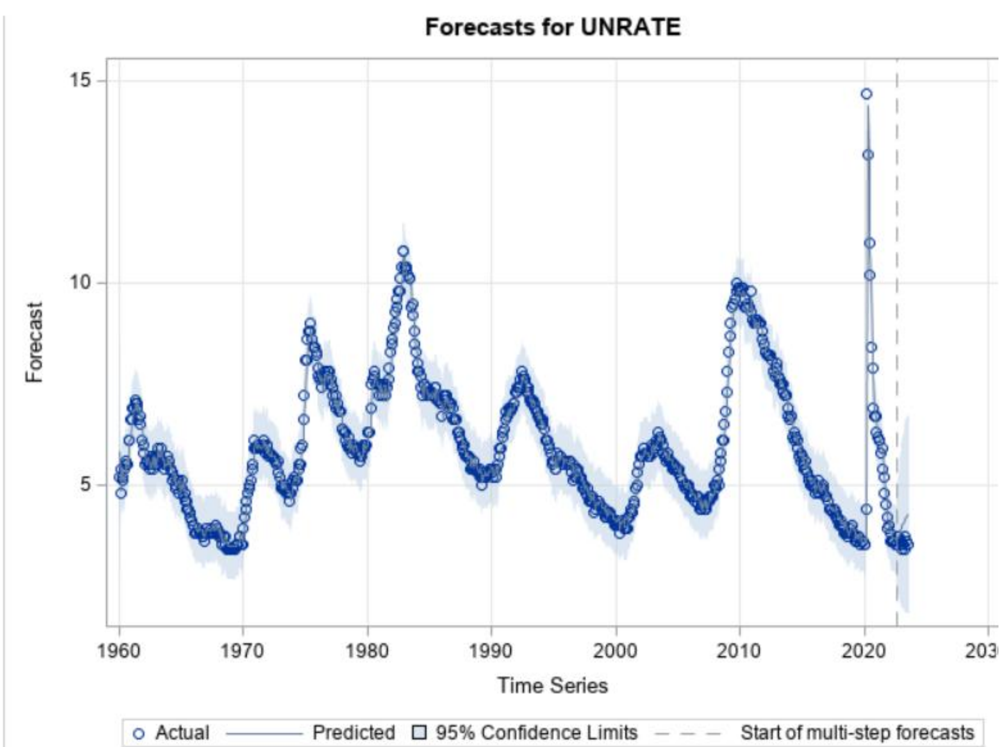


Fig 5.22 Forecasting Results

Fit Statistics:

Total rows: 11 Total columns: 3

	TYPE	_STAT_	_VALUE_
1	ML	AIC	883.76127847
2	ML	SBC	893.03579453

Fig:5.23 Fit Statistics

Using the Macros to derive the accuracy for the Model.

```
/* STSM03s04c.sas */


%let nhold=12;
%include "C:\Study\Sem-2(Fall)\Data Mining\forecasting\Forecasting-datasets-OPIM5671\Macros2.sas" / source2;
%accuracy_prep(indsn=stsm.'DM FC PRJ'n, series=UNRATE, timeid='Time Series'n,
    numholdback=&nhold);
```

```
proc arima data=WORK._TEMP plots=none;
    identify var=_y_fit;
    estimate p=(1) method=ML;
    forecast lead=&nhold id='Time Series'n interval=month out=work.outARMa1 nooutall;
    run;
```

```
%accuracy(indsn=WORK.outARMa1, timeid='Time Series'n, series=UNRATE,
    numholdback=&nhold);
```

Series	Model	Holdback Periods	MAPE	MAE	MSE	RMSE
UNRATE	WORK.outARMa1	12	11.90%	0.41909	0.23250	0.48218

Fig: 5.24 Accuracy Macro

Achieving 88.10% accuracy using ARMA (1,0,0) model.

5.4 Model 4: - ARMAX:

In the model 4 ARMAX we have added independent variables for the model prediction. Those Independent variables are Job openings and the federal interest rate.

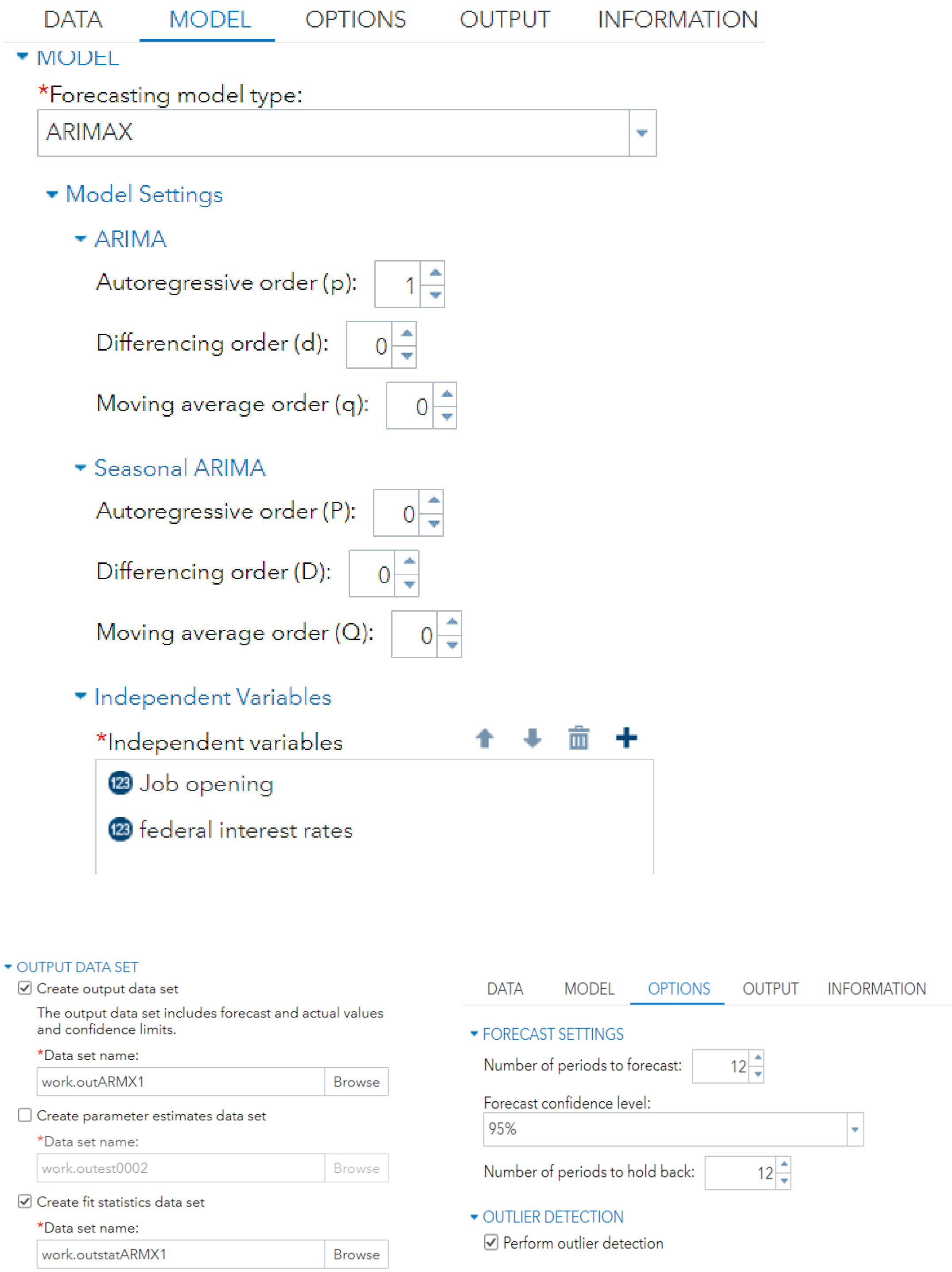


Fig: 5.25 Model fitting

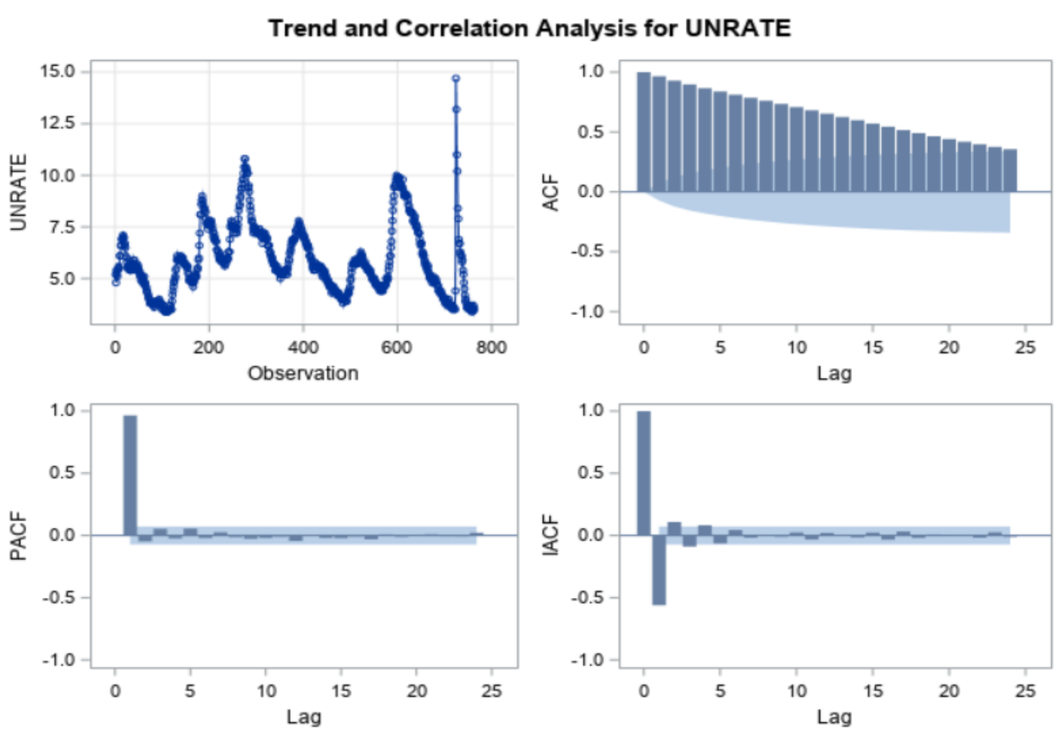
Model Results:

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	9.87476	0.60656	16.28	<.0001	0	UNRATE	0
AR1,1	0.96748	0.0091026	106.29	<.0001	1	UNRATE	0
NUM1	-0.0007298	0.00008138	-8.97	<.0001	0	Job opening	0
NUM2	-0.09521	0.02887	-3.30	0.0010	0	federal interest rates	0

Constant Estimate	0.321114
Variance Estimate	0.165294
Std Error Estimate	0.406563
AIC	798.6151
SBC	817.1641
Number of Residuals	763

Fig 5.26 Maximum Likelihood Estimates

Here in this observation, we can understand that mean (MU) & AR1,1 and the 2 exogenous variables (Job Opening, Federal Interest Rate) are significant.



Correlations of Parameter Estimates				
Variable Parameter	UNRATE MU	UNRATE AR1,1	Job opening NUM1	federal interest rates NUM2
UNRATE MU	1.000	0.102	-0.658	-0.211
UNRATE AR1,1	0.102	1.000	-0.146	0.020
Job opening NUM1	-0.658	-0.146	1.000	-0.024
federal interest rates NUM2	-0.211	0.020	-0.024	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	12.43	5	0.0293	-0.040	-0.049	0.042	-0.100	-0.006	-0.017
12	18.98	11	0.0615	0.004	-0.001	0.022	-0.005	0.088	0.011
18	21.80	17	0.1925	0.034	0.012	0.028	0.029	-0.003	0.026
24	23.43	23	0.4361	-0.011	0.018	-0.005	0.025	0.024	-0.020
30	27.68	29	0.5351	0.005	-0.027	0.036	-0.051	0.024	0.013
36	30.88	35	0.6674	0.007	0.045	-0.038	0.013	-0.010	0.014
42	33.36	41	0.7962	-0.011	-0.007	-0.048	0.004	0.002	-0.024
48	36.42	47	0.8677	0.021	0.002	-0.014	0.035	-0.044	0.005

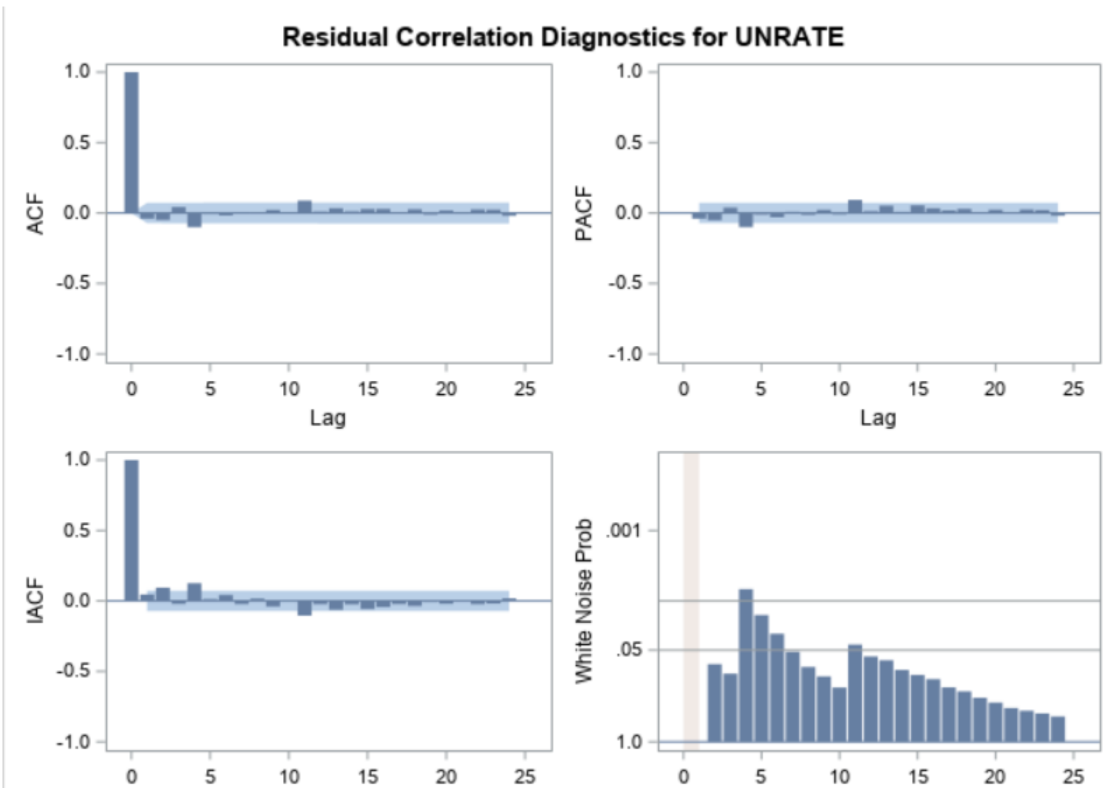


Fig:5.27 Residual Correlation

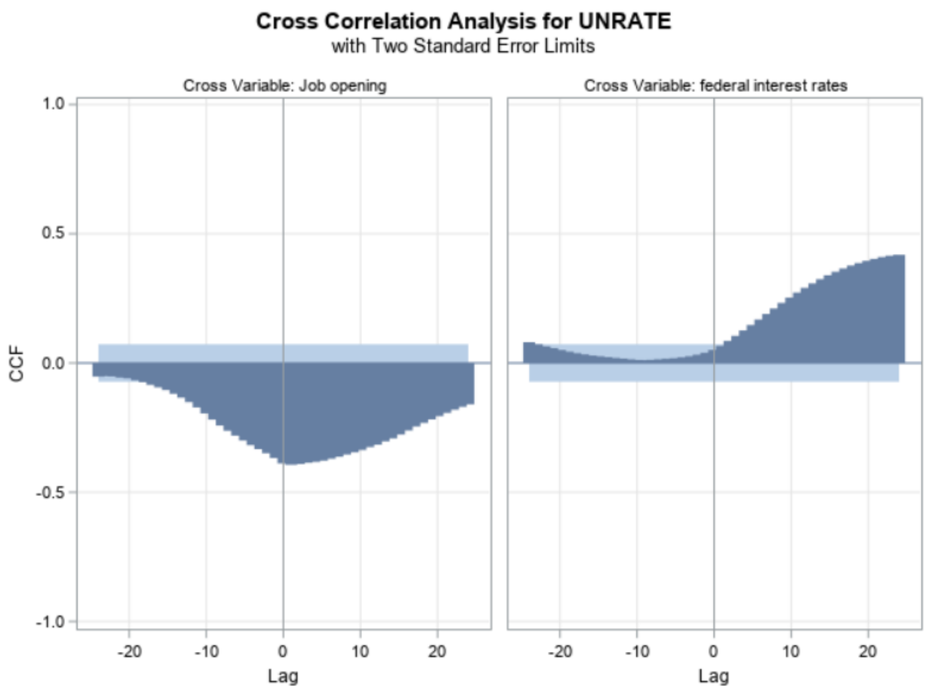


Fig 5.28 White Noise test and Cross Correlation Observation

The white noise probability test suggests the presence of some signal in the error. Furthermore, the cross-correlation between the variables resembles a time series pattern. To mitigate this, pre-whitening both variables is necessary.

5.5 Pre whiten techniques:

Job Opening:

After applying the pre whiten technique to the job openings feature the cross-correlation unemployment rate and the job openings are

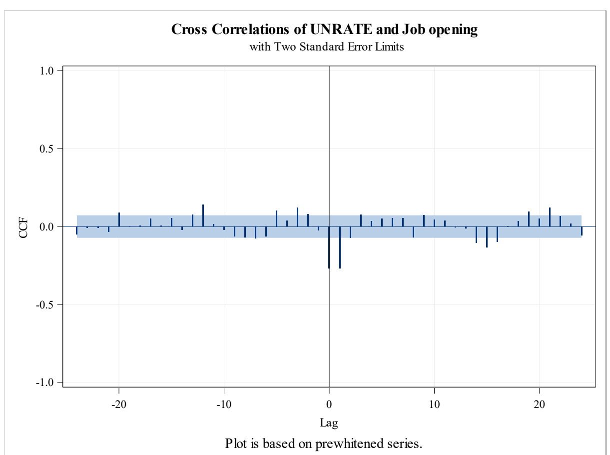


Fig 5.29 Cross Correlations for Job Openings

From Above Graph we see that the cross-correlation of target variable UNRATE with input Job opening is Significant at the Lag 1.

Residual plots with white noise test results express that there is no signal in the data.

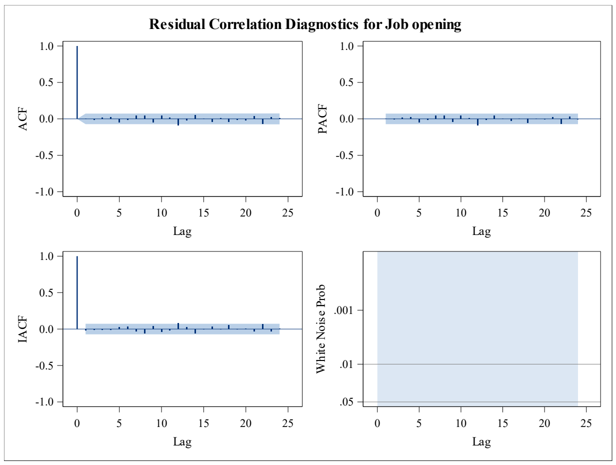


Fig 5.30 Residuals for Job openings for unemployment rate

Federal Interest Rate:

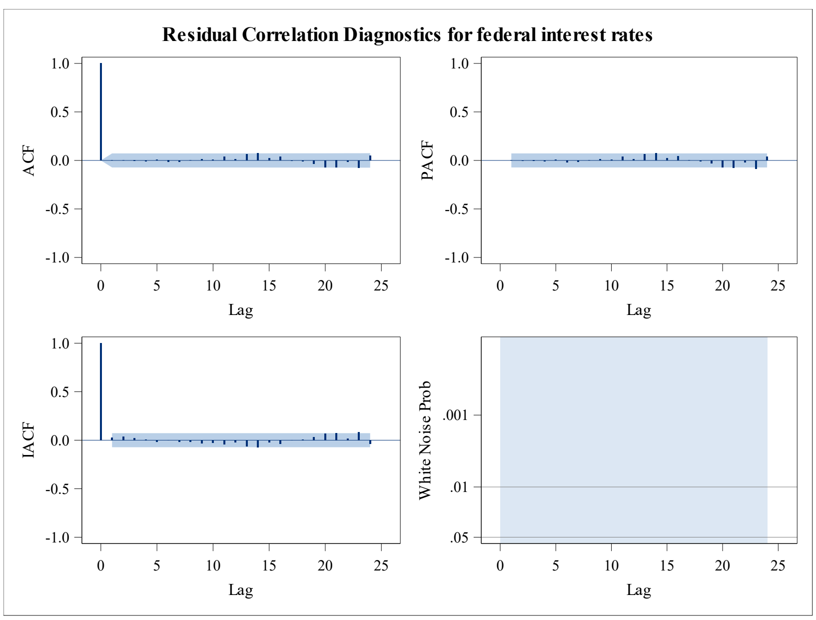


Fig 5.31 Federal Interest rate residuals

The Residual correlation of the federal interest rate after applying pre whitening in the white noise test there is no signal in the data.

The cross correlation of the unemployment rate and the Federal Interest Rate

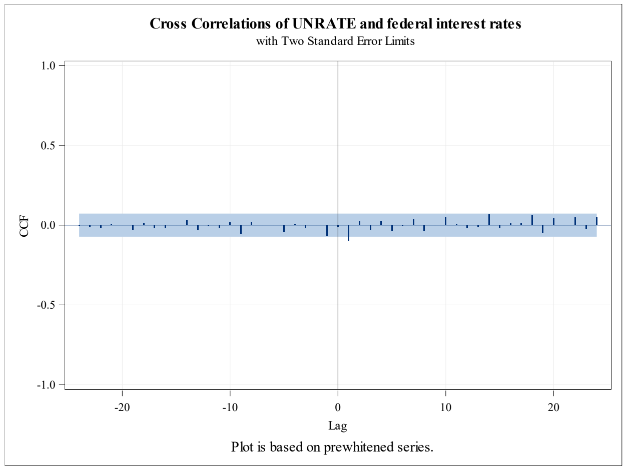


Fig 5.32 Cross Correlations for Interest Rate

Also, with the above Graph we could say that the cross-correlation of target variable UNRATE with input federal interest rate is Significant at the Lag 1.

By implementing pre-whitening techniques on the Exogeneous variables, we've effectively addressed autocorrelation issues, resulting in a model that adeptly captures and predicts data patterns at significant lags.

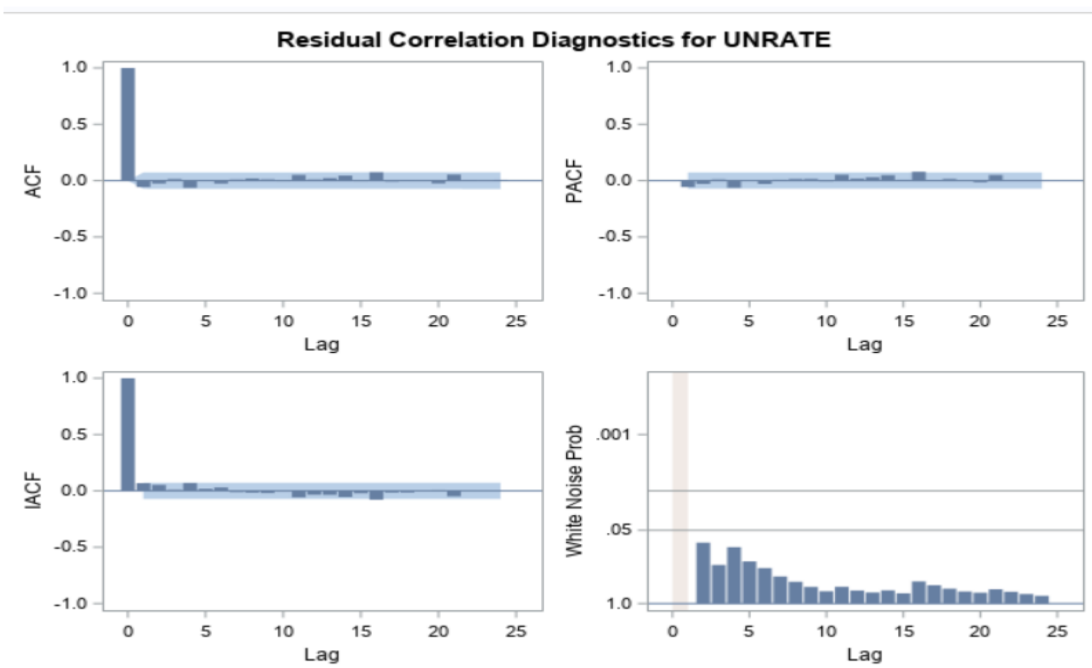


Fig 5.33 Residual Analysis

Upon scrutinizing the residual plot post Pre-whitening, followed by the adjustment of significant lag effects in the code and subsequent model rerun, it becomes evident that the noise has significantly decreased compared to the pre-whitening stage. This reduction in noise proves instrumental in enhancing model development and overall accuracy.

Forecasting:

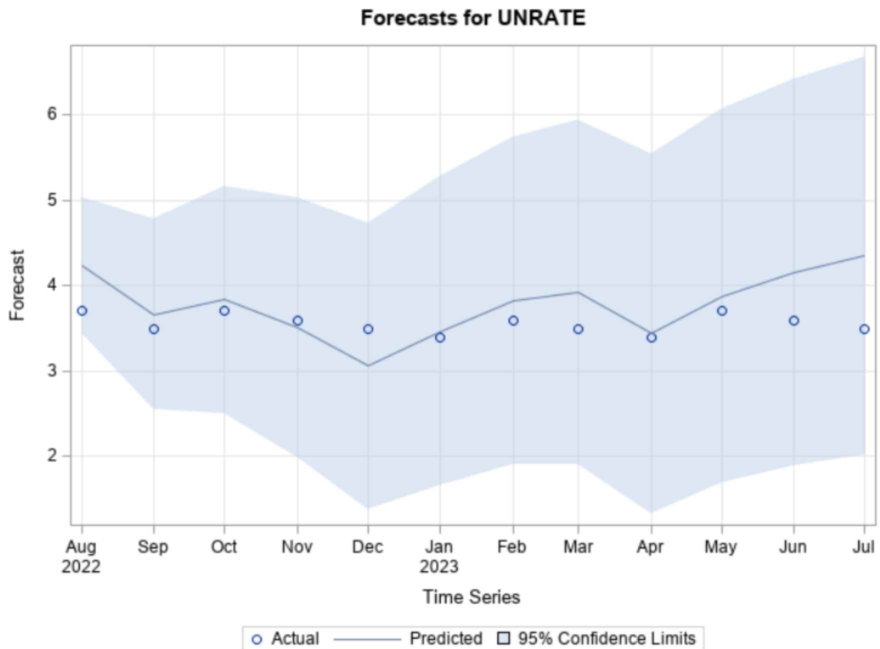
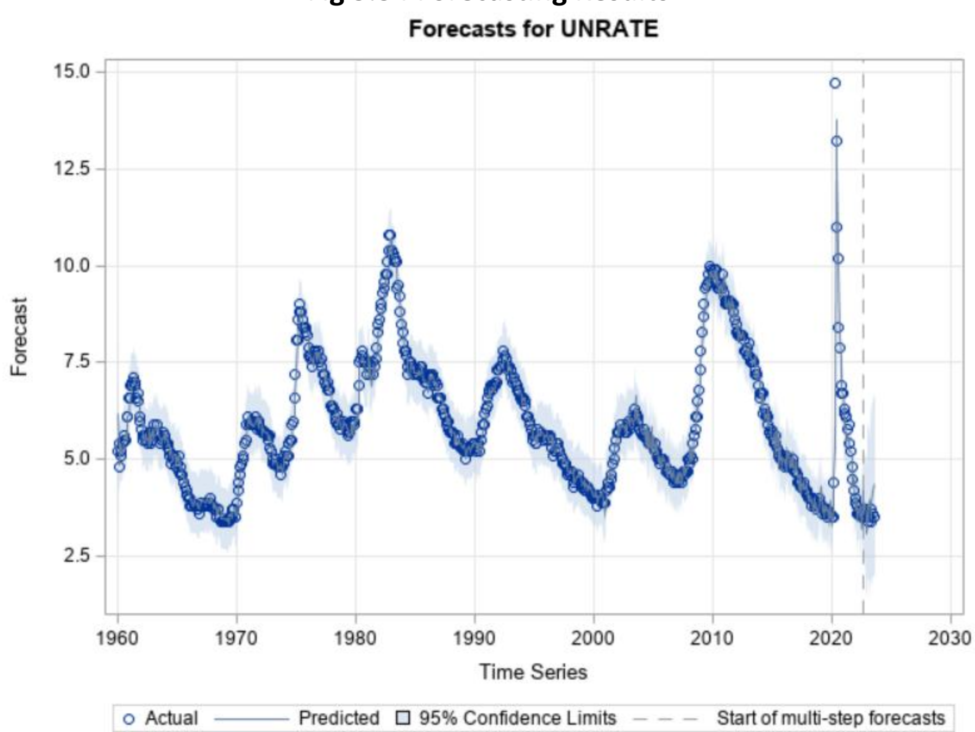


Fig 5.34 Forecasting Results



Fit Statistics:

Total rows: 11 Total columns: 3

	TYPE	_STAT_	_VALUE_
1	ML	AIC	831.34279094
2	ML	SBC	849.88657716

Fig 5.35 Fit Statistics

Using the Macros to derive the accuracy for the Model.

```
/* STSM03s04c.sas */


%let nhold=12;
%include "C:\Study\Sem-2(Fall)\Data Mining\forecasting\Forecasting-datasets-OPIM5671\Macros2.sas" / source2;
%accuracy_prep(indsn=stsm.'DM FC PRJ'n, series=UNRATE, timeid='Time Series'n,
    numholdback=&nhold);

```

```
/* STSM03s04d.sas */
ods select none;

proc arima data=WORK._TEMP plots=none;
    identify var=_y_fit crosscorr=('Job opening'n 'federal interest rates'n);;
    estimate p=(1) input=(1 $ 'Job opening'n 1 $ 'federal interest rates'n) method=ML;
    forecast lead=&nhold id='Time Series'n interval=month out=work.outARMX1 nooutall;
    run;
```

```
%accuracy(indsn=work.outARMX1, timeid='Time Series'n, series=UNRATE,
    numholdback=&nhold);
```

Total rows: 1 Total columns: 7 Rows 1-1

	Model	Series	Holdback Periods	MAPE	MAE	MSE	RMSE
1	work.outARMX1	UNRATE	12	9.55%	0.3400795655	0.1514677893	0.3891886295

Fig 5.36 Accuracy Macro

Achieving 90.45% accuracy using ARMAX (1,0,0) model.

5.6 Model Comparison:

The following are fit statistics of the three models.

Model	Accuracy	MAPE	AIC	SBC
ARMA (2,0,0)	87.05%	12.95%	884.02	897.93
ARMA (1,0,0)	88.10%	11.90%	883.76	893.03
ARIMA (1,1,0)	97.15%	2.85%	891.73	901
ARMAX (1,0,0)	90.45%	9.55%	831.34	849.88

Fig 5.37 Model Comparison

Choosing the best model for training the data:
After comparing all three models we would prefer the ARIMA(1,1,0) model as the best model. We have come to this conclusion considering the Accuracy and MAPE error values. The accuracy of the ARIMA(1,1,0) model is 97.15%. It is better to choose a model which is cost-effective and robust.

6. Recommendations:

6.1 Implications for unemployment:

- By Emphasizing the effectiveness of incorporating job openings and the federal interest rate as exogenous (independent) variables in your model. These variables are relevant predictors of unemployment rates and contribute to the model's forecasting performance.
- The robustness of the pre whitening technique in enhancing the quality of our independent variables. By removing autocorrelation and making them stationary, pre whitening ensures that the predictors are more suitable for modeling.
- The accurate forecasts generated by your ARMAX model can offer valuable insights to policymakers and economic analysts. These forecasts can aid in evidence-based decision-making and policy formulation, especially during economic fluctuations.
- Businesses and investors can benefit from the model's predictions. Accurate unemployment rate forecasts can guide hiring decisions, workforce planning, and investment strategies, contributing to more informed and profitable choices.

6.2 Limitations and Challenges:

- As a team, we encountered challenges related to limited historical data and gaps in the dataset, which impacted the accuracy of our long-term forecasts. We also had to contend with data quality issues, including missing values and measurement errors, which required careful handling.

- We recognized that the ARIMAX model assumes a linear relationship between independent variables (job openings and federal interest rate) and the dependent variable (unemployment rate). As a team, we were mindful that deviations from linearity could affect the model's predictive ability.
- External economic events, policy changes, and unforeseen factors posed challenges in our forecasting efforts. As a team, we acknowledged that such external influences, which may not be accounted for in the model, could significantly impact unemployment rates.
- Handling the complexity of ARIMAX models, especially when incorporating multiple independent variables, was a challenge for our team. Balancing model complexity to avoid overfitting or issues with model interpretation required careful consideration.
- Ensuring the stationarity of time series data after differencing was a critical step for our team. Recognizing the significance of this assumption, we worked together to validate data stationarity.

6.3 Business Insights:

- Unemployment rate predictions generated by the ARIMAX model offer policymakers valuable insights for informed decision-making. Accurate forecasts enable policymakers to formulate targeted labor market policies, such as job training programs and unemployment benefit adjustments, to address changing economic conditions effectively.
- The ARIMAX model's accuracy in predicting unemployment rates provides governments and organizations with the ability to allocate resources efficiently. It allows for precise planning of workforce development initiatives and investment strategies based on expected labor market conditions.
- Financial institutions and investors can benefit from the ARIMAX model's precise forecasts when assessing potential risks associated with lending, investments, and portfolio management. Accurate predictions assist in evaluating the impact of labor market dynamics on financial stability.

7. References:

1. https://data.oecd.org/unemp/unemployment-rate-forecast.htm

2. https://tradingeconomics.com/united-states/unemployment-rate

3. https://howmuch.net/articles/timeline-united-states-unemployment-history