Time Series Analysis in Forecasting United States Teenage Female Unemployment

Sekyiwaah Nuamah[†][*]

[†]*Department of Mathematics, East Tennessee State University, Johnson City, USA.*

Abstract

This study analyzes monthly unemployment patterns for United States (U.S.) females aged 16-19 from January 1950 to April 2025 using SAS time series procedures. The research addresses critical socioeconomic needs to understand the dynamics of youth unemployment, particularly among teenage females who face unique challenges in the labor market, including limited work experience, seasonal employment patterns, and systemic inequalities. This demographic historically experiences unemployment rates exceeding 15%, with increased risk during economic downturns like the 2008 financial crisis and COVID-19 pandemic. Using PROC ARIMA, we checked for the stationarity of the time plot, identify model through ACF/PACF analysis, and SARIMA model fitting. After evaluating multiple candidate models, $SARIMA(0,1,1)(0,1,1)_{12}$ emerged as the optimal specification based on AIC criteria and diagnostic checks. The model successfully captured seasonal patterns and short-term dynamics, generating reliable 12-month forecasts with 95% confidence intervals revealing expected seasonal fluctuations consistent with school-year employment patterns. Findings provide actionable insights for policymakers developing targeted youth employment programs and workforce development strategies.

1 Introduction

Having a job as a teenager can be a valuable experience, teaching responsibility, and organizational and time management skills, along with providing a paycheck [1].

Unemployment according to U.S. Bureau of Labor Statistics (BLS) are persons who were not employed during the reference week, are available for work during that week and had made specific efforts to find work in the four-week period ending in the reference week [2]. Unemployment remains one of the most important indicators of the economic health and social well-being of a country. It affects not only national productivity but also individual livelihoods and long-term opportunities. Among various groups affected by unemployment, young people, particularly females between the ages of 16 and 19, are among the most vulnerable.

In the United States, the teenage labor force is especially sensitive to seasonal changes, school schedules, and shifts in the economy. Teens who do in fact want jobs face competition from older workers, young college graduates, and foreign-born workers [8]. Young females face additional challenges due to gender-based disparities, childcare responsibilities, and a higher likelihood of working in part-time or low-wage service-sector jobs [3–5]. During economic downturns such as the 2008 financial crisis or the COVID-19 pandemic in 2020, young people are often among the first to lose their jobs and among the last to recover. According to the U.S. Bureau of Labor Statistics (BLS), the unemployment rate for females aged 16 to 19 has historically been higher than the national average, and tends to fluctuate more dramatically across months and years [6, 7].

In an era where data drives decisions, understanding how teenage women participate in the workforce is more than just a statistic, it is a window into the progress of the nation. Therefore, understanding the patterns in youth unemployment, especially in terms of trends and seasonality, is important for developing targeted policies and support systems. Accurately forecasting future unemployment rates can help government agencies, educational institutions, and community organizations plan interventions, training programs, and job placement services more effectively. It can also guide the allocation of resources in areas such as education, social services, and workforce development.

[*]nuamah@etsu.edu

2 Purpose of the research

The central focus of this research is understanding how the rate of teenage female unemployment in the U.S. has changed month to month, and whether those changes follow any predictable trends and use these to forecast future employment level.

The purpose of this study is to

1. To identify the best-fitting statistical model,
2. To forecast future unemployment patterns, and
3. To generate meaningful insights into the economic realities faced by teenage females in the U.S. labor market.

3 Description of Data

The data were obtained from the U.S. Bureau of Labor Statistics (BLS) through the Current Population Survey (CPS), a nationally representative household survey conducted jointly by the U.S. Census Bureau and the BLS.

The data were collected using standardized labor force survey instruments where respondents self-report their employment status based on definitions aligned with international labor statistics standards.

This study uses monthly time series data on the number of unemployed females aged 16 to 19 in the United States. The dataset spans multiple decades, with each row representing a single month from the period of January 1950 to April 2025, providing over 900 monthly observations as shown in figure 6.1. The dataset used in this project was downloaded in CSV format

The variables used in the study are Date, which is in the form of month and year, and Value which represents the number of unemployed teenage females (ages 16–19) in thousands for each month. This is the dependent variable in the forecast model (type: continuous numeric). The data are measured monthly and are seasonally unadjusted, they naturally include recurring seasonal fluctuations offering a more authentic view of unemployment dynamics.

This time-ordered structure of the dataset makes it suitable for applying time series techniques such as differencing, autocorrelation analysis, and ARIMA/SARIMA modeling, all aimed at identifying underlying trends and generating forecasts for future months.

4 Proposed Analyses

To address the research questions, this project will follow a structured time series analysis workflow using SAS statistical software. The overall goal is to model the underlying patterns in the monthly employment data of females in the U.S. and produce accurate short-term forecasts.

The analysis involves both descriptive exploration and inferential model fitting. Our first step involves visualizing the raw time series data to examine overall trends and detect any apparent shifts or patterns and if there is variance instability, I first stabilize variance to achieve a more linear trend, by applying a log transformation of the Value variable. This transformation will help satisfy the assumptions of many time series models, particularly regarding homoscedasticity and normality of residuals. The next thing we will do is to test the series for stationarity using the Augmented Dickey-Fuller (ADF) test via proc arima. Stationarity is a key assumption for fitting ARIMA models. If the ADF test indicates non-stationarity, the data is difference firstly with the First-order differencing to remove trend and Seasonal differencing (lag 12) to address any remaining structure. We then identify and estimate the model based on autocorrelation (ACF) and partial autocorrelation (PACF) plots, as well as fit statistics (AIC, SBC).

We then check if the model is adequate through residual autocorrelation tests (Ljung-Box Q-statistics), residual plots, normality plot like histogram and Q-Q plot and also check for independence using Durbin Wastin test. Once model adequacy is confirmed, we finally generate forecasts for the next 12 months. These are initially on the log scale the forecasts are accompanied by 95% confidence intervals, enabling interpretation of prediction uncertainty. The final forecast result is visualized using proc sgplot and finally we check the prediction using the RMSE, MSE.

5 Analysis

5.1 Model Specification and Model Fitting

5.1.1 Time Plot and Preliminary Observations

A time plot of the monthly unemployment rate for teenage females in the U.S. was generated using `PROC SGPLOT` to visually assess the structure of the data. From the plot in figure 6.2, it was observed that the variance appeared relatively stable, but there was clear evidence of non-stationarity due to a recurring seasonal cycles and visible cyclical patterns and irregular variation(spikes).There was also pattern of trend but not as clear as seasonal or cyclic, its weak relative to seasonal and cyclical effects. this was confirmed by the R^2 value of the linear trend test in figure 6.3. These patterns suggested that further statistical testing and transformation were necessary before proceeding with modeling.

5.1.2 Stationarity Check and Differencing

To formally assess whether the series was stationary, a Augmented Dickey-Fuller test was conducted which results showed that the series is stationary under most specifications ($p < 0.05$) depicted in figure 6.5. However, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots in figure 6.4 were examined. The slow decay in the ACF plot and the strong autocorrelation at seasonal lags confirmed that the original series was non-stationary. The PACF plot also depict non-stationarity

To address this, a two-step differencing approach was applied. They were First-order differencing ($d = 1$) to remove the trend shown in figure 6.6 and Seasonal differencing ($D = 1$) at lag 12 to eliminate yearly seasonality shown in figure 6.7.

After applying the combined differencing, `Value(1,12)`, the transformed series showed no remaining trend, seasonal correlation, cyclic or irregularities, and the ACF/PACF plots indicated stationarity in figure 6.7.

5.1.3 Model Identification and Candidate Models

Several Seasonal ARIMA ($SARIMA(p, d, q)(P, D, Q)_S$) models were specified and tested to identify the best-fitting model. (p,d,q) represent the non seasonal and $(P, D, Q)_s$ represent the seasonal. The model selection process considered both Statistical criteria such as the Akaike Information Criterion (AIC), SBC, Variance estimate, P-value and Model diagnostics including residual normality, white noise assumptions, and independence.

The following SARIMA configurations were evaluated: $SARIMA(0, 1, 1)(0, 1, 1)_{12}$, $SARIMA(1, 1, 0)(1, 1, 0)_{12}$, $SARIMA(0, 1, 1)(1, 1, 1)_{12}$ and $SARIMA(0, 1, 0)(0, 1, 1)_{12}$,

The models are shown in figure 6.8, 6.9, 6.10, 6.11

5.2 Model Fitting

Each candidate model was estimated using Maximum Likelihood Estimation (MLE) through `PROC ARIMA` in SAS.

The model $SARIMA(0, 1, 1)(0, 1, 1)_{12}$ emerged as the best-performing specification based on the lowest AIC value of 3511.024 and SBC of 3525.401, and the lowest variance estimate value of 2.961772. The model has fewer parameters making it a simpler structure for estimation and efficiency. The model also has well-behaved residuals. All these results are shown in figure 6.8

The estimation results showed that both the non-seasonal MA(1) and the seasonal MA(1) terms were statistically significant ($p < 0.05$), confirming their contribution to capturing the short-term and seasonal behavior in the series.

5.3 Model Diagnostics

To validate the adequacy of the final model, several diagnostic checks were performed

1. Residual ACF and PACF plots confirmed that no significant autocorrelation remained as shown in figure 6.12 .

2. The Ljung-Box Q graphically showed that the residuals resembled white noise ($p > 0.05$). However, there were some autocorrelation among residuals statistically as shown in figure 6.14. The residual autocorrelation diagnostics showed only small autocorrelation values at lower lags, and visual inspection confirmed randomness, supporting the adequacy of the model .

3. Normality of residuals was assessed using histogram and Q-Q plots, which showed approximate normal distribution with a few minor outliersas shown in figure 6.13. However, formal goodness-of-fit tests (Kolmogorov–Smirnov, Cramér–von Mises, and Anderson–Darling) rejected the null hypothesis of normality at the 5% level as shown in figure 6.16. This deviation is likely due to the extreme spike in unemployment during the COVID-19 period.

4. Independence of residuals was visually confirmed by randomness in the residual plots and statistically through Durbin Watson test which produced values close to 2, indicating little to no serial correlation as shown in figure 6.12, 6.15.

These diagnostics confirmed that the $SARIMA(0,1,1)(0,1,1)_{12}$ model was statistically adequate and suitable for producing reliable forecasts.

5.4 Forecasting Results

After the $SARIMA(0,1,1)(0,1,1)_{12}$ model was confirmed to be statistically adequate through residual diagnostics, it was used to generate a 12-month ahead forecast of the monthly unemployment rate for U.S. teenage females. The forecast was implemented using the `FORECAST` statement in `PROC ARIMA` in SAS.

5.4.1 Forecast Output

The forecast table shows the predicted unemployment rates, with their corresponding standard errors, and the 95% confidence intervals for each month in the forecast period. Predicted values range from approximately 8.38% to 13.34%, with initial increase then a decline occurring from 13.34% to 8.38% of the forecast period before a mild increase then back to a decrease toward the end. These changes align with the historical seasonal cycle for this demographic, where unemployment typically declines during the school year and increases slightly during the summer months. The standard errors and confidence intervals widen over time, reflecting increasing uncertainty as the forecast horizon extends. Figure 6.17 represents the forecast results.

5.4.2 Forecast Plot

The forecast plot illustrates the forecasted unemployment rates along with their corresponding 95% confidence intervals. The forecast pattern suggests a mild decline in unemployment in the early forecast periods, followed by slight upward and downward fluctuations that are consistent with historical seasonal effects observed in participation in teenage labor. These short-term oscillations(fluctuation) reflect expected school-to-summer cycles common in this demographic. As expected, the confidence bands widen as the forecast horizon extends, reflecting greater uncertainty in long-term predictions in figure 6.18.

Prediction Statistics

The forecasting model for the unemployment rate of teenage females achieved a Mean Squared Error (MSE) of 2.99 and a Root Mean Squared Error (RMSE) of 1.73 based on 916 observations. This indicates that, on average, the model prediction deviate from the actual unemployment rate by approximately 1.73 percentage points. The relatively low RMSE suggests that the model provides a reasonably accurate fit, making it useful for capturing the dynamics of teenage female unemployment. This is shown in figure 6.19.

6 Conclusion

The purpose of this study was to analyze and identify an appropriate time series model for teenage female unemployment in the United States, and to forecast future unemployment patterns under the assumption that existing labor and welfare policies would remain stable.

The analysis began with an exploration of the historical data, which revealed seasonality, cyclic, irregularity, and trend in the unemployment rate of teenage girls in the U.S. These patterns were confirmed through visual plots and autocorrelation analysis. To make the data suitable for forecasting, first-order and seasonal differencing were applied, and several SARIMA models were tested and evaluated.

After comparing models, $SARIMA(0, 1, 1)(0, 1, 1)_{12}$ was identified as the most suitable, based on statistical criteria and diagnostic checks. This model effectively captured the repeating seasonal patterns and short-term changes in the data.

The resulting 12-month forecast suggests modest fluctuations in the unemployment rate over the next year, with seasonal dips and rises consistent with school-year and summer employment patterns. These results highlight the real-world challenges that teenage girls face in finding steady work, as their employment often depends on seasonal and part-time opportunities. The forecast can help guide leaders in planning youth employment programs, education policies and initiatives, and labor market policy and interventions. By predicting future unemployment patterns for teenage females, decision makers can better align support services, training programs, and outreach strategies to meet the needs of young women in the labor force.

APPENDIX

DATA

https://data.bls.gov/timeseries/LNU04000014
Source:Bureau of Labor Statistics (BLS) Unemployment Rate for Teenage Females from 16 to 19 years(not seasonally adjusted).
Data accessed and analyzed January 1950 through April 2025

SAS CODES

```
/*  Import the data (assumes CSV format) */
proc import datafile="/home/u63993191/Unemployment.csv"
    out=unemployment
    dbms=csv
    replace;
    getnames=yes;
run;

/*  Create SAS-formatted monthly date starting from Jan 1948 */
data unemployment;
    set unemployment;
    date = intnx('month', '01JAN1950'd, _n_-1);
    format date monyy7.;
run;

proc print data=unemployment;
run;

/*Keeping the columns to use*/
data unemployment;
    set unemployment(keep=Value);
    date = intnx('month', '01JAN1950'd, _n_-1);
    format date monyy7.;
run;

proc print data=unemployment;
run;

/*  Plot time series */
proc sgplot data=unemployment;
    series x=date y=Value / lineattrs=(thickness=2 color=blue);
    xaxis label="Time";
    yaxis label="Unemployment Rate (%)";
    title "Monthly U.S. Unemployment Rate";
run;
*check for stationary;
proc arima data=unemployment;
    identify var=Value stationarity=(adf);
run;
data unemployment;
    set unemployment(keep=Value);
    date = intnx('month', '01JAN1950'd, _n_-1);
    time_index = _n_;  /* Create sequential time index */
    format date monyy7.;
run;
```

```
/*  Linear Regression Trend Test */
proc reg data=unemployment;
   model Value = time_index;
   title "Linear Trend Test - Regression";
run;
* To remove trend, we start with differencing of order 1(d=1);
*fIRST ORDER DIFFERENCING;
proc arima data=unemployment;
identify var=Value(1) stationarity=(adf);
run;

* DIFFERENCING OF SEASONAL COMPONENT;
proc arima data=unemployment;
identify var=Value(1, 12) stationarity=(adf);
run;

 * Model fitting - SARIMA(0,1,1)(0,1,1);
proc arima data=unemployment;
identify var=Value(1,12) stationarity=(adf);
estimate q=(1)(12) method=ML;
forecast lead=12 out=forecast_unemployment;
Title "SARIMA (0,1,1)(0,1,1)";
run;

 * Model fitting - SARIMA(0,1,1)(1,1,1);
proc arima data=unemployment;
identify var=Value(1,12) stationarity=(adf);
estimate p=(12) q=(1)(12) method=ML;
forecast lead=12 out=forecast_unemployment;
Title "SARIMA (0,1,1)(1,1,1)";
run;

* Model fitting - SARIMA(0,1,0)(0,1,1);
proc arima data=unemployment;
identify var=Value(1,12) stationarity=(adf);
estimate  q=(12) method=ML;
forecast lead=12 out=forecast_unemployment;
Title "SARIMA (0,1,0)(0,1,1)";
run;

 * Model fitting - SARIMA(0,1,0)(1,1,1);
proc arima data=unemployment;
identify var=Value(1,12) stationarity=(adf);
estimate p=(12) q=(12) method=ML;
forecast lead=12 out=forecast_unemployment;
Title "SARIMA (0,1,1)(1,1,1)";
run;

 * Model fitting - SARIMA(0,1,1)(2,1,1);
proc arima data=unemployment;
identify var=Value(1,12) stationarity=(adf);
estimate p=(24) q=(1)(12) method=ML;
forecast lead=12 out=forecast_unemployment;
Title "SARIMA (0,1,1)(2,1,1)";
run;
```

```
 * Model fitting - SARIMA(1,1,0)(1,1,0);
proc arima data=unemployment;
identify var=Value(1,12) stationarity=(adf);
estimate p=(1)(12) method=ML;
forecast lead=12 out=forecast_unemployment;
Title "SARIMA (1,1,0)(1,1,0)";
run;

** Using this model SARIMA(0,1,1)(0,1,1);
proc arima data=unemployment;
identify var=Value(1,12) stationarity=(adf);
estimate q=(1)(12) method=ML;
forecast lead=12 out=forecast_unemployment;
Title "SARIMA (0,1,1)(0,1,1)";
run;
* DURBIN-WATSON TEST;
proc autoreg data=forecast_unemployment;
model residual=/DW=3 DWprob  ; run;
* TEST OF NORMALITY;
proc univariate data=forecast_unemployment;
var residual;
histogram/cfill=blue normal(color=red);
run;
* Prediction;
data forecast_eval;
    set forecast_unemployment;
    if not missing(Value) and not missing(Forecast) then do;
        Error = Value - Forecast;
        SqError = Error**2;
    end;
run;

proc means data=forecast_eval noprint;
    var SqError;
    output out=eval_stats mean=MSE;
run;

data eval_stats;
    set eval_stats;
    RMSE = sqrt(MSE);
run;

proc print data=eval_stats label;
    label MSE = "Mean Squared Error"
          RMSE = "Root Mean Squared Error";
    Title "Prediction";
run;
```

SAS OUTPUTS

A snapshot of the data

Figure 6.1

Obs	Value	date
1	12.1	JAN1950
2	13.3	FEB1950
3	13.1	MAR1950
4	8.2	APR1950
5	11.2	MAY1950
6	17.8	JUN1950
7	14.5	JUL1950
8	9.1	AUG1950
9	10.4	SEP1950
10	7.7	OCT1950
11	8.9	NOV1950
12	9.9	DEC1950
13	8.1	JAN1951
14	8.3	FEB1951
15	8.6	MAR1951
16	7.2	APR1951
17	6.5	MAY1951
18	13.6	JUN1951
19	10	JUL1951
20	7.4	AUG1951
21	8.6	SEP1951
22	5.6	OCT1951
23	8	NOV1951
24	5.8	DEC1951
25	8.5	JAN1952
26	6.8	FEB1952
27	6.6	MAR1952
28	5.6	APR1952
29	9.2	MAY1952

Figure 6.1: A sample of the data.

Time plot

Figure 6.2

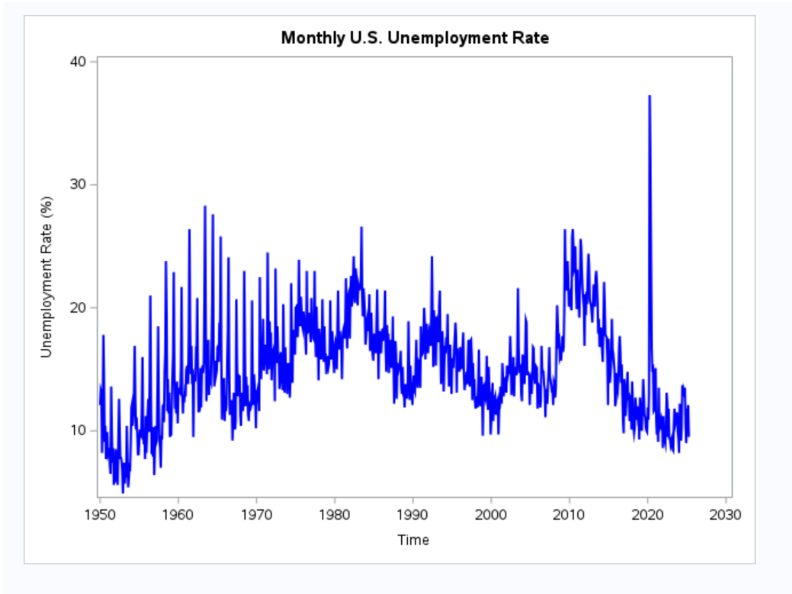

Figure 6.2: Time plot of unemployment rate.

Stationary checks

Figure 6.3

Linear Trend Test - Regression

The REG Procedure
Model: MODEL1
Dependent Variable: Value

Number of Observations Read	904
Number of Observations Used	904

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	345.11389	345.11389	20.72	<.0001
Error	902	15026	16.65879		
Corrected Total	903	15371			

Root MSE	4.08152	R-Square	0.0225
Dependent Mean	15.10188	Adj R-Sq	0.0214
Coeff Var	27.02656		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	14.03051	0.27172	51.64	<.0001
time_index	1	0.00237	0.00052019	4.55	<.0001

Figure 6.3: Trend test confirming non-stationarity.

Figure 6.4

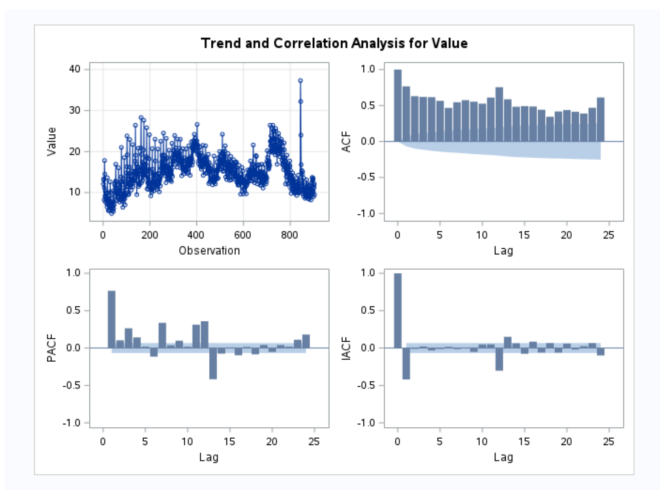

Figure 6.4: ACF and PACF plots used for stationarity check.

Figure 6.5

The ARIMA Procedure

Name of Variable = Value	
Mean of Working Series	15.10188
Standard Deviation	4.123555
Number of Observations	904

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2076.58	6	<.0001	0.766	0.630	0.620	0.616	0.565	0.467
12	4049.52	12	<.0001	0.546	0.574	0.553	0.526	0.611	0.755
18	5311.01	18	<.0001	0.587	0.483	0.491	0.486	0.441	0.344
24	6500.64	24	<.0001	0.417	0.437	0.414	0.389	0.469	0.612

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-14.6965	0.0076	-2.73	0.0063		
	1	-9.8491	0.0292	-2.23	0.0247		
	2	-5.0373	0.1229	-1.60	0.1031		
Single Mean	0	-209.852	0.0001	-10.86	<.0001	59.02	0.0010
	1	-170.280	0.0001	-9.19	<.0001	42.21	0.0010
	2	-94.4976	0.0018	-6.66	<.0001	22.16	0.0010
Trend	0	-214.306	0.0001	-10.97	<.0001	60.21	0.0010
	1	-174.757	0.0001	-9.28	<.0001	43.12	0.0010
	2	-96.9733	0.0008	-6.71	<.0001	22.56	0.0010

Figure 6.5: ADF test.

Differencing

Figure 6.6

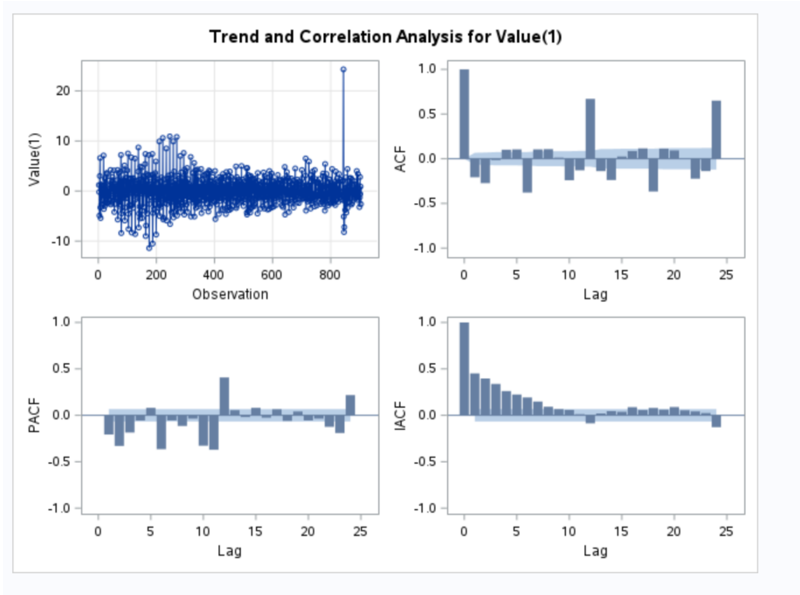

Figure 6.6: first order differencing

Figure 6.7

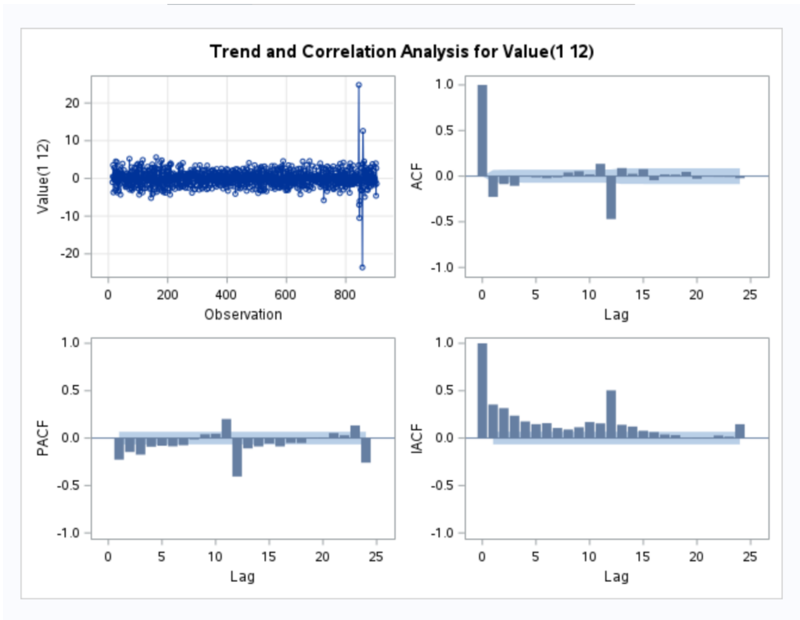

Figure 6.7: seasonal differencing.

Model Identification

Figure 6.8

SARIMA (0,1,1) (0,1,1)

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.0001965	0.0080977	0.02	0.9806	0
MA1,1	0.35254	0.03126	11.28	<.0001	1
MA2,1	0.79335	0.02143	37.03	<.0001	12

Constant Estimate	0.000196
Variance Estimate	2.961772
Std Error Estimate	1.72098
AIC	3511.024
SBC	3525.401
Number of Residuals	891

Correlations of Parameter Estimates			
Parameter	MU	MA1,1	MA2,1
MU	1.000	0.004	-0.005
MA1,1	0.004	1.000	0.026
MA2,1	-0.005	0.026	1.000

Figure 6.8: SARIMA (0,1,1)(0,1,1).

Figure 6.9

SARIMA(0,1,1)(1,1,1)

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.0001389	0.0077484	0.02	0.9857	0
MA1,1	0.35313	0.03130	11.28	<.0001	1
MA2,1	0.80998	0.02592	31.25	<.0001	12
AR1,1	0.03421	0.04241	0.81	0.4199	12

Constant Estimate	0.000134
Variance Estimate	2.962555
Std Error Estimate	1.721207
AIC	3512.492
SBC	3531.661
Number of Residuals	891

Correlations of Parameter Estimates				
Parameter	MU	MA1,1	MA2,1	AR1,1
MU	1.000	0.004	-0.003	0.002
MA1,1	0.004	1.000	0.047	0.041
MA2,1	-0.003	0.047	1.000	0.602
AR1,1	0.002	0.041	0.602	1.000

Figure 6.9: SARIMA (0,1,1)(1,1,1).

Figure 6.10

SARIMA (0,1,0)(0,1,1)

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.0002149	0.01275	-0.02	0.9865	0
MA1,1	0.79690	0.02130	37.42	<.0001	12

Constant Estimate	-0.00021
Variance Estimate	3.166674
Std Error Estimate	1.779515
AIC	3569.681
SBC	3579.266
Number of Residuals	891

Correlations of Parameter Estimates		
Parameter	MU	MA1,1
MU	1.000	-0.001
MA1,1	-0.001	1.000

Figure 6.10: SARIMA (0,1,0)(0,1,1).

Figure 6.11

(1,1,0)(1,1,0)

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.0039782	0.03641	0.11	0.9130	0
AR1,1	-0.21731	0.03276	-6.63	<.0001	1
AR2,1	-0.47181	0.02962	-15.93	<.0001	12

Constant Estimate	0.007127
Variance Estimate	3.770723
Std Error Estimate	1.941835
AIC	3717.208
SBC	3731.585
Number of Residuals	891

Correlations of Parameter Estimates			
Parameter	MU	AR1,1	AR2,1
MU	1.000	-0.002	0.000
AR1,1	-0.002	1.000	-0.038
AR2,1	0.000	-0.038	1.000

Figure 6.11: SARIMA (1,1,0)(1,1,0).

Model Adequacy Check

Figure 6.12

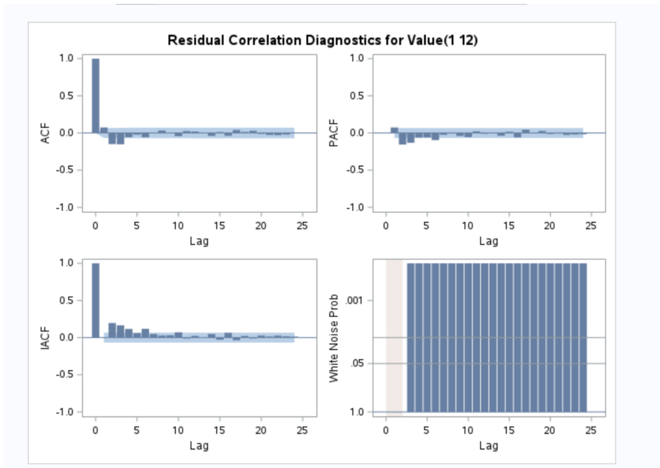

Figure 6.12: residual plots.

Figure 6.13

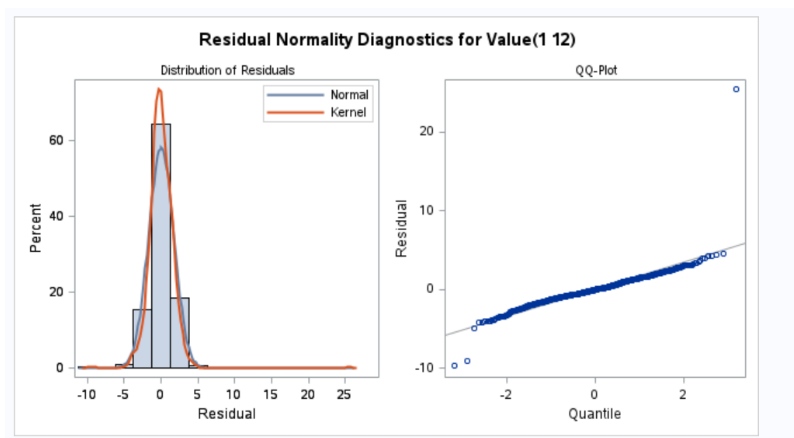

Figure 6.13: normality plots.

Figure 6.14

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	51.88	4	<.0001	0.074	-0.149	-0.152	-0.057	-0.022	-0.058
12	55.43	10	<.0001	-0.001	0.033	-0.004	-0.042	0.026	0.020
18	59.84	16	<.0001	0.006	-0.039	0.012	-0.037	0.040	0.014
24	62.49	22	<.0001	0.029	-0.011	-0.026	-0.029	-0.021	-0.003
30	66.18	28	<.0001	0.040	0.003	0.002	0.001	0.049	0.002
36	72.35	34	0.0001	-0.052	-0.031	-0.043	0.011	-0.030	0.013
42	75.42	40	0.0006	0.017	0.022	0.013	0.000	-0.042	-0.024
48	79.47	46	0.0016	0.045	-0.024	-0.011	0.015	-0.035	0.008

Figure 6.14: autocorrelation.

Figure 6.15

Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	1.8554	0.0153	0.9847
2	2.2897	1.0000	<.0001
3	2.2863	1.0000	<.0001

Figure 6.15: independence tests.

Figure 6.16

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.06882688	Pr > D	<0.010
Cramer-von Mises	W-Sq	1.47941696	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	9.81249407	Pr > A-Sq	<0.005

Figure 6.16: Normality test

Forecasting

Figure 6.17

Forecasts for variable Value				
Obs	Forecast	Std Error	95% Confidence Limits	
905	11.5943	1.7210	8.2212	14.9674
906	13.3442	2.0502	9.3258	17.3625
907	11.4188	2.3334	6.8454	15.9923
908	10.6642	2.5858	5.5961	15.7324
909	10.9141	2.8157	5.3955	16.4328
910	10.3870	3.0281	4.4519	16.3221
911	9.2658	3.2266	2.9417	15.5899
912	8.3825	3.4136	1.6919	15.0730
913	10.0771	3.5909	3.0391	17.1151
914	9.7447	3.7598	2.3757	17.1138
915	9.3108	3.9214	1.6250	16.9967
916	10.1559	4.0767	2.1657	18.1460

Figure 6.17: forecast output.

Figure 6.18

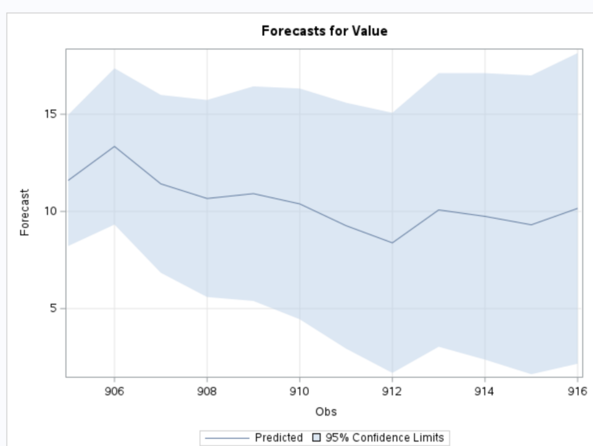

Figure 6.18: forecast plot.

Prediction Statistics

Figure 6.19

Prediction

Obs	_TYPE_	_FREQ_	Mean Squared Error	Root Mean Squared Error
1	0	916	2.99263	1.72992

Figure 6.19: prediction accuracy.

References

[1] Morisi, Teresa L. "Youth enrollment and employment during the school year." Monthly Lab. Rev. 131 (2008): 51.

[2] Levitan, Mark. Unemployment and Joblessness in New York City, 2006: Recovery Bypasses Youth. Community Service Society of New York, February 2007. https://smhttp-ssl-58547.nexcesscdn.net/nycss/images/uploads/pubs/UnemploymentandJoblessnessinNYC2006.PDF

Levitan, Mark. Unemployment and Joblessness in New York City, 2005. Community Service Society of New York, 2004. https://www.cssny.org/publications/entry/unemployment-and-joblessness-in-new-york-city-2005

[3] Price, Brenda M., and Michael Wasserman. "Gender Gaps in the Labor Market Widen Every Summer." Board of Governors of the Federal Reserve System, March 24, 2023. https://www.federalreserve.gov/econres/notes/feds-notes/gender-gaps-in-the-labor-market-widen-every-summer-20230324.html.

[4] Cortés, Patricia, and Jessica Pan. Children and the Remaining Gender Gaps in the Labor Market. NBER Working Paper No. 27980. Cambridge, MA: National Bureau of Economic Research, 2020. https://www.nber.org/system/files/working_papers/w27980/w27980.pdf.

[5] U.S. Bureau of Labor Statistics. "Employment and Unemployment among Youth–Summer 2024." U.S. Department of Labor, August 20, 2024. https://www.bls.gov/news.release/youth.nr0.htm.

[6] Congressional Research Service. Unemployment Rates During the COVID-19 Pandemic: In Brief. Updated February 15, 2021. https://www.congress.gov/crs_external_products/R/PDF/R46554/R46554.7.pdf

[7] U.S. Bureau of Labor Statistics. A-10. Unemployment Rates by Age, Sex, and Marital Status, Seasonally Adjusted. Last modified August 1, 2025. https://www.bls.gov/web/empsit/cpseea10.htm

[8] Morisi, Teresa L. "Teen labor force participation before and after the Great Recession and beyond." Monthly Lab. Rev. 140 (2017): 1.