Q4

1. Introduction

In this question, I fit a **Seasonal ARIMA (SARIMA) model** to the **unemployment data** from the `astsa` package.
The goals are:

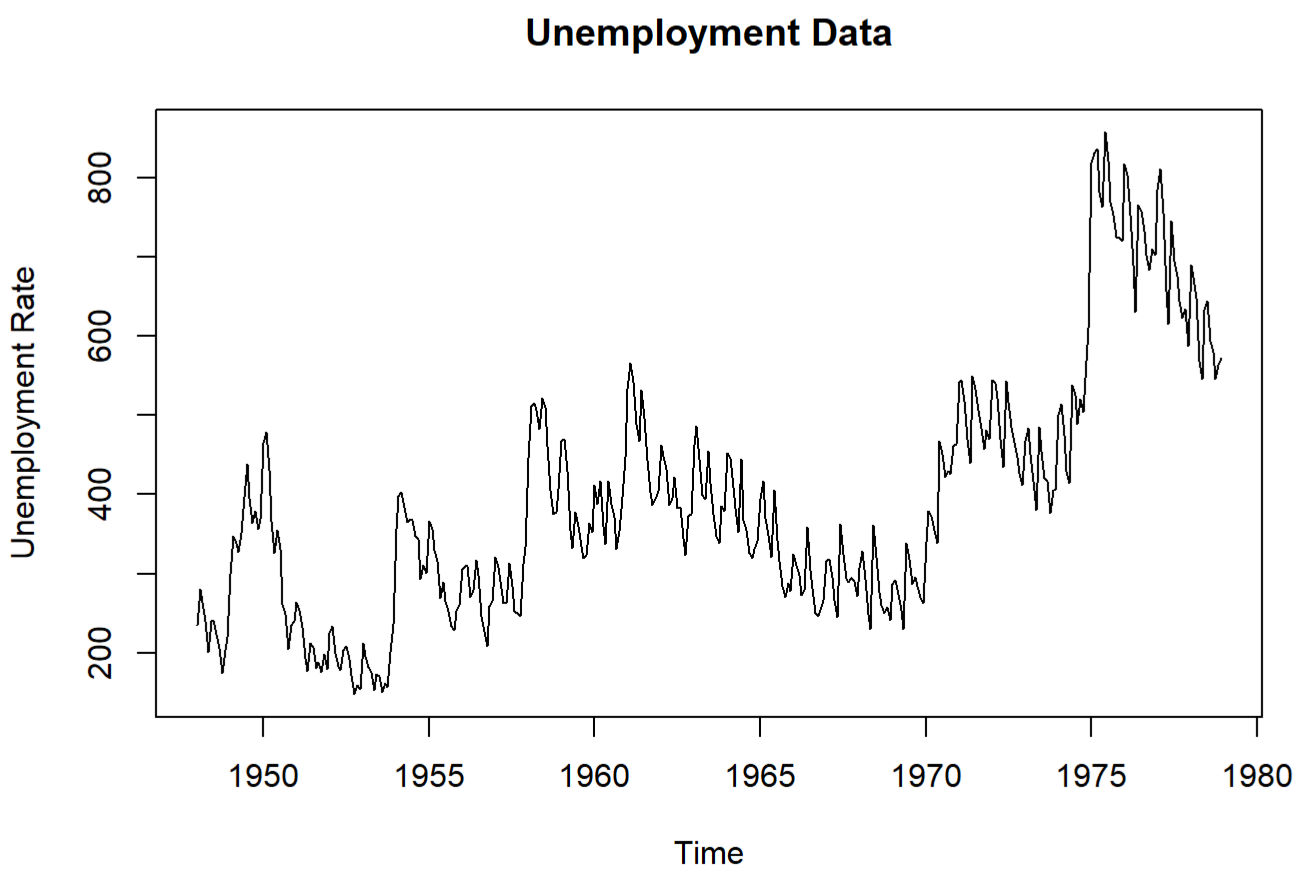
1. Estimate an appropriate **SARIMA model**.
2. Forecast unemployment for the **next 12 months**.
3. Provide detailed model diagnostics and report findings properly.

2. Load Libraries and Data

```
# Load necessary libraries
library(forecast)
library(tseries)
library(astsa)

# Load the unemployment data
data("unemp")

# Plot the original data to visualize trends and seasonality
plot(unemp, main = "Unemployment Data", ylab = "Unemployment Rate", xlab = "Time")
```



3. Observations from ACF/PACF and Model Selection

After plotting the unemployment data x_t, it is clear that:

1. **Seasonality** is present with a yearly pattern, which requires seasonal differencing.
2. **ACF and PACF Insights**:
    - **ACF**: Seasonal lags cut off after **lag 12** (lags 12, 24, 36…).
    - **PACF**: Seasonal lags tail off slowly after lags 12, 24, 36.

From these observations: - Fit a **SARIMA(0, 1, 0) × (0, 1, 1)_{12}** model to x_t and analyze the ACF and PACF of residuals. - A seasonal AR model (like AR(2)) is suggested for within-season analysis or **ARMA(1,1)** for short-term dependencies.

Both SARIMA(2, 1, 0) × (0, 1, 1)*{12} (model (i)) and SARIMA(1, 1, 1) × (0, 1, 1)*{12} (model (ii)) have the same number of parameters.
Model (i) is preferred due to **lower MSE** and **whiter residuals**.

---

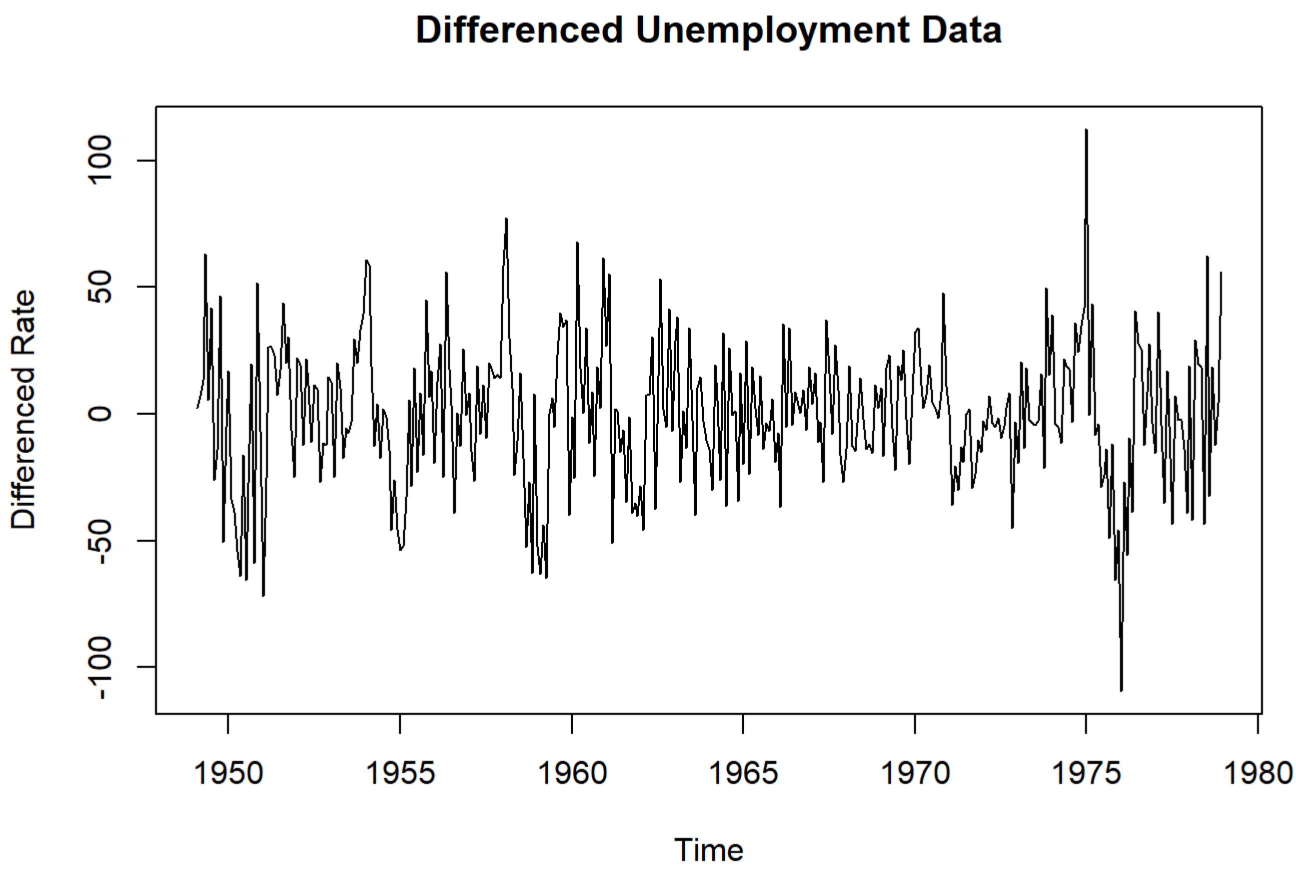
4. Differencing to Achieve Stationarity

The unemployment data requires both **seasonal** and **non-seasonal** differencing to achieve stationarity:

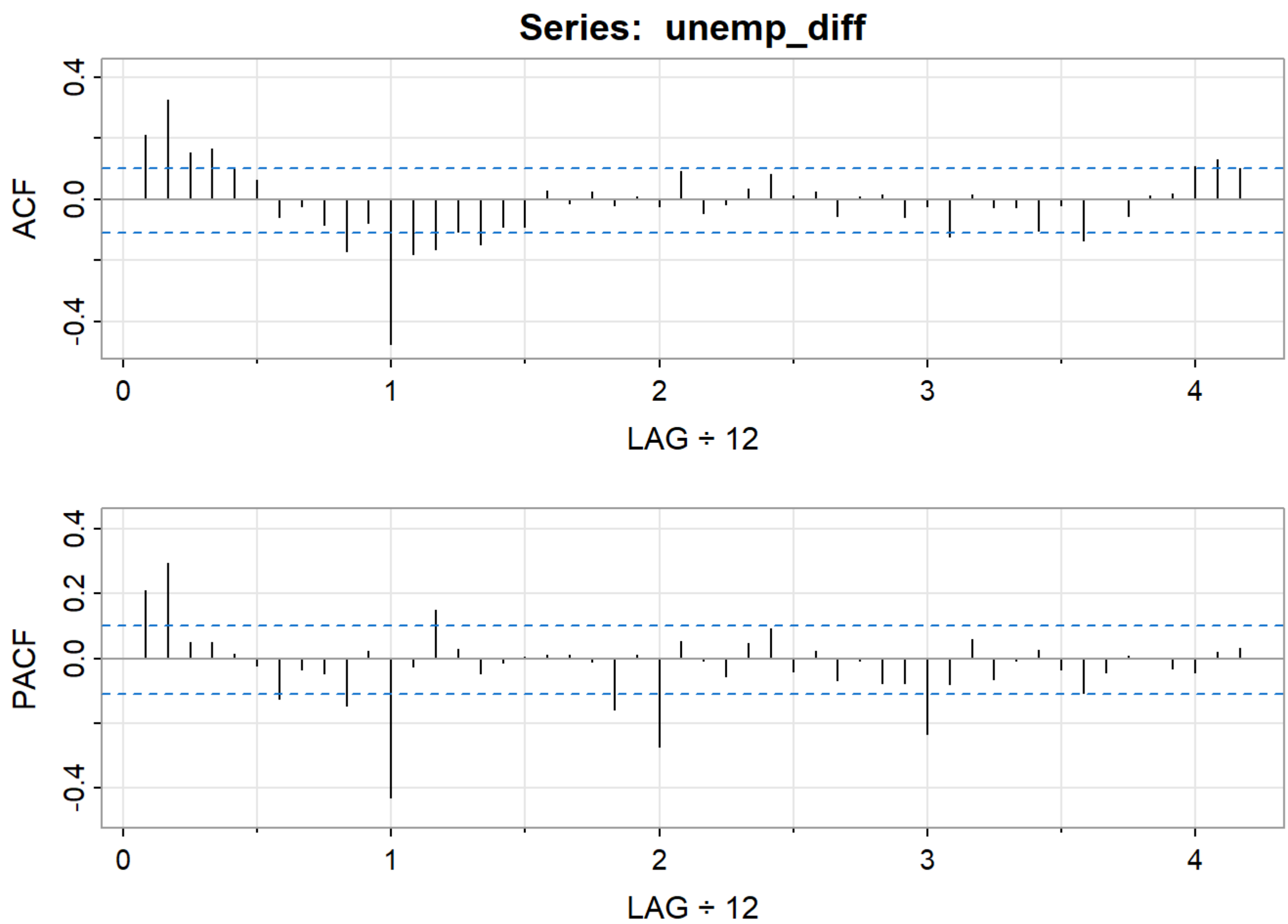
$$\nabla_{12} y_t = y_t - y_{t-12}, \quad \nabla y_t = y_t - y_{t-1}$$

```
# Apply seasonal and non-seasonal differencing
unemp_diff <- diff(diff(unemp, lag = 12))

# Plot the differenced data
plot(unemp_diff, main = "Differenced Unemployment Data", ylab = "Differenced Rate")
```




```
# ACF and PACF of the differenced data
acf2(unemp_diff, 50)
```

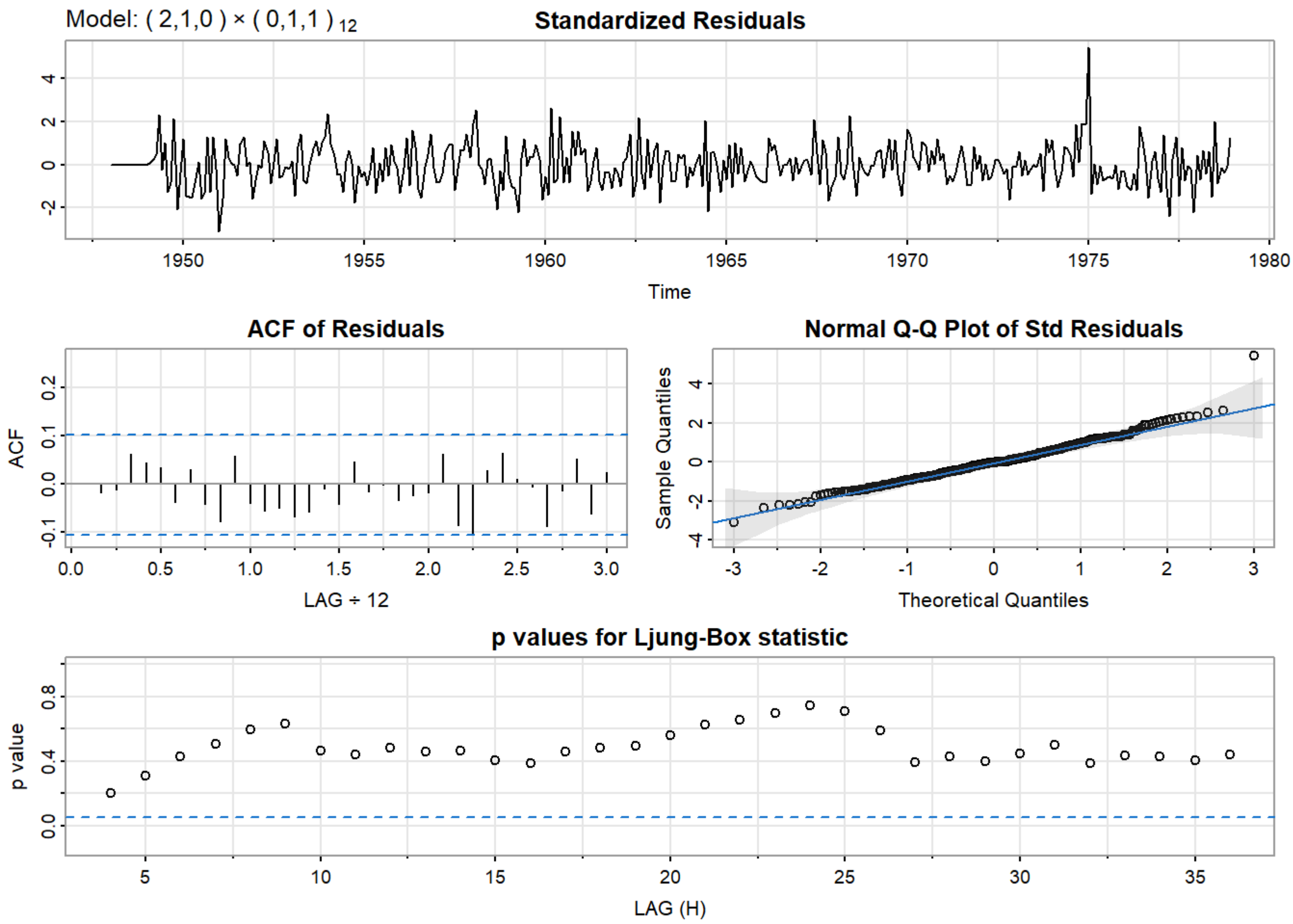


```
##      [,1] [,2] [,3] [,4] [,5]  [,6]  [,7]  [,8]  [,9] [,10] [,11] [,12] [,13]
## ACF  0.21 0.33 0.15 0.17 0.10  0.06 -0.06 -0.02 -0.09 -0.17 -0.08 -0.48 -0.18
## PACF 0.21 0.29 0.05 0.05 0.01 -0.02 -0.12 -0.03 -0.05 -0.15  0.02 -0.43 -0.02
##      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25]
## ACF  -0.16 -0.11 -0.15 -0.09 -0.09  0.03 -0.01  0.02 -0.02  0.01 -0.02  0.09
## PACF  0.15  0.03 -0.04 -0.01  0.00  0.01  0.01 -0.01 -0.16  0.01 -0.27  0.05
##      [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35] [,36] [,37]
## ACF  -0.05 -0.01  0.03  0.08  0.01  0.03 -0.05  0.01  0.02 -0.06 -0.02 -0.12
## PACF -0.01 -0.05  0.05  0.09 -0.04  0.02 -0.07 -0.01 -0.08 -0.08 -0.23 -0.08
##      [,38] [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46] [,47] [,48] [,49]
## ACF   0.01 -0.03 -0.03 -0.10 -0.02 -0.13  0.00 -0.06  0.01  0.02  0.11  0.13
## PACF  0.06 -0.07 -0.01  0.03 -0.03 -0.11 -0.04  0.01  0.00 -0.03 -0.04  0.02
##      [,50]
## ACF   0.10
## PACF  0.03
```

5. Fitting SARIMA Models

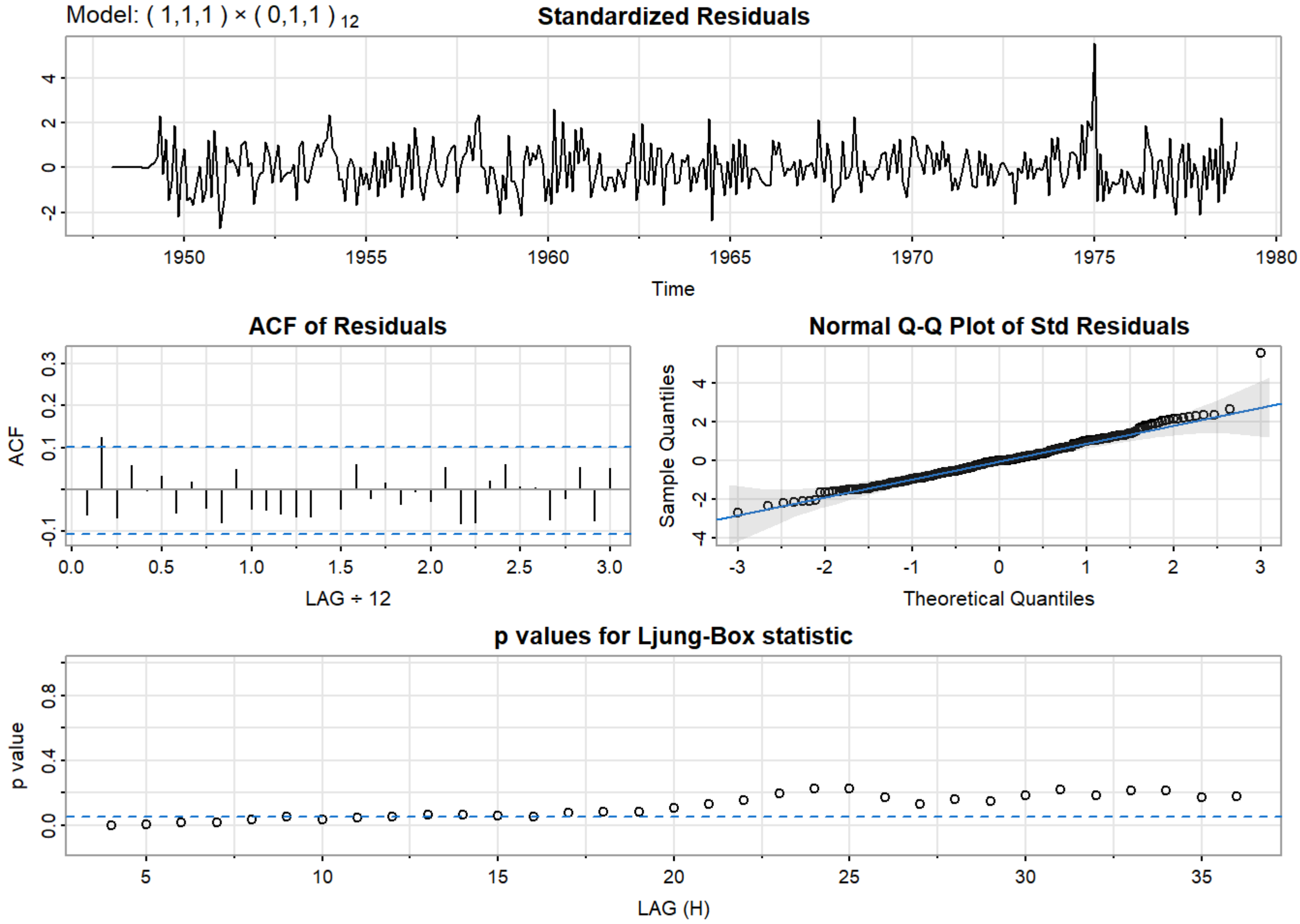
```
# Fit SARIMA(2, 1, 0) × (0, 1, 1)[12] model (preferred model (i))
sarima_model_1 <- sarima(unemp, 2, 1, 0, 0, 1, 1, 12)
```

```
## initial  value 3.340809 
## iter   2 value 3.105512
## iter   3 value 3.086631
## iter   4 value 3.079778
## iter   5 value 3.069447
## iter   6 value 3.067659
## iter   7 value 3.067426
## iter   8 value 3.067418
## iter   8 value 3.067418
## final  value 3.067418 
## converged
## initial  value 3.065481 
## iter   2 value 3.065478
## iter   3 value 3.065477
## iter   3 value 3.065477
## iter   3 value 3.065477
## final  value 3.065477 
## converged
## <><><><><><><><><><><><><><>
## 
## Coefficients: 
##      Estimate     SE  t.value p.value
## ar1    0.1351 0.0513   2.6326  0.0088
## ar2    0.2464 0.0515   4.7795  0.0000
## sma1  -0.6953 0.0381 -18.2362  0.0000
## 
## sigma^2 estimated as 449.637 on 356 degrees of freedom 
##  
## AIC = 8.991114  AICc = 8.991303  BIC = 9.034383 
## 
```




```
# Fit SARIMA(1, 1, 1) × (0, 1, 1)[12] model (alternative model (ii))
sarima_model_2 <- sarima(unemp, 1, 1, 1, 0, 1, 1, 12)
```

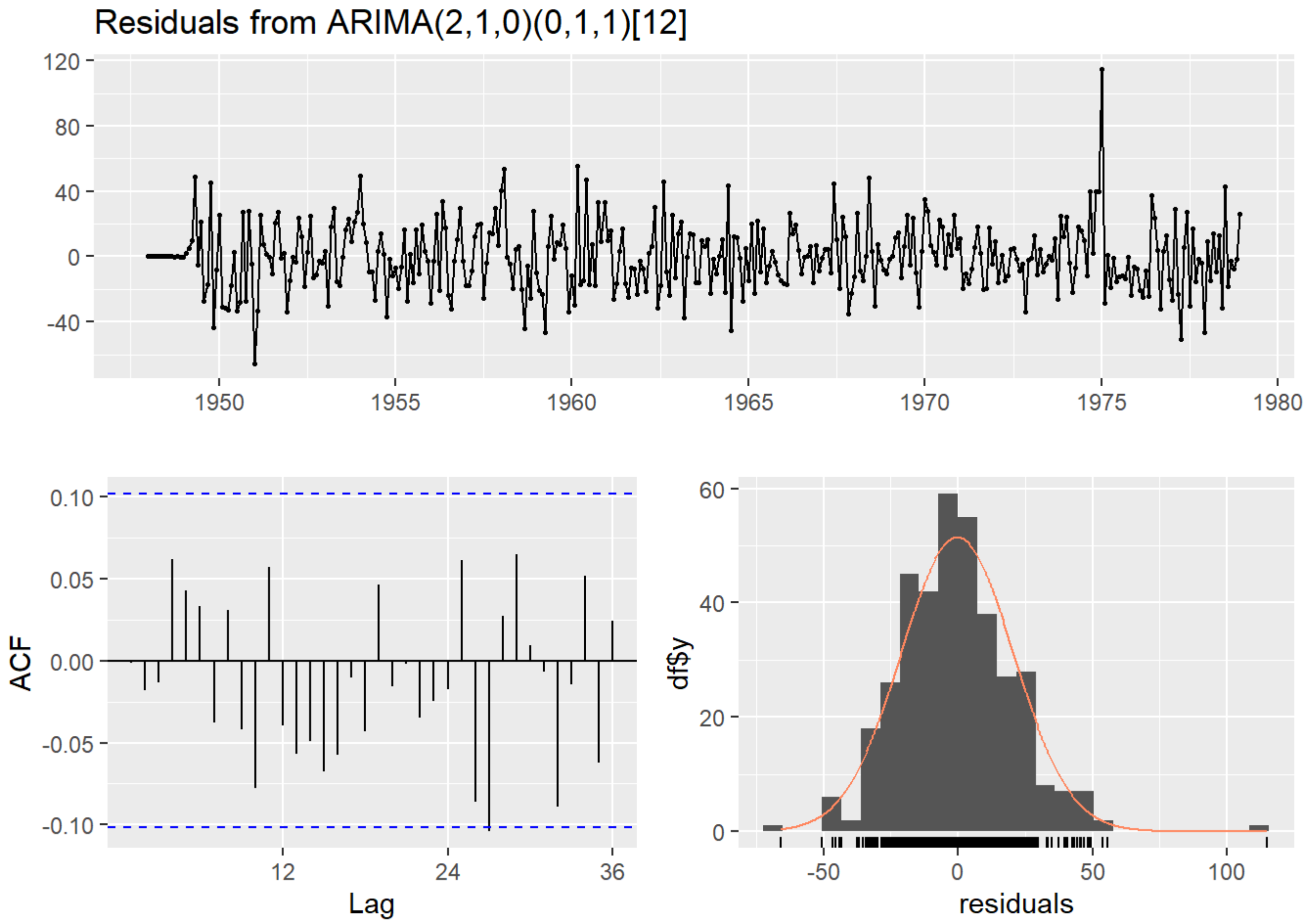
```
## initial  value 3.339497 
## iter   2 value 3.185867
## iter   3 value 3.137572
## iter   4 value 3.103005
## iter   5 value 3.102720
## iter   6 value 3.096553
## iter   7 value 3.095284
## iter   8 value 3.093866
## iter   9 value 3.093270
## iter  10 value 3.091924
## iter  11 value 3.084816
## iter  12 value 3.078887
## iter  13 value 3.076774
## iter  14 value 3.075302
## iter  15 value 3.075003
## iter  16 value 3.074831
## iter  17 value 3.074802
## iter  18 value 3.074783
## iter  19 value 3.074770
## iter  20 value 3.074768
## iter  21 value 3.074767
## iter  21 value 3.074767
## iter  21 value 3.074767
## final  value 3.074767 
## converged
## initial  value 3.074581 
## iter   2 value 3.074580
## iter   3 value 3.074578
## iter   4 value 3.074578
## iter   5 value 3.074577
## iter   6 value 3.074577
## iter   6 value 3.074577
## iter   6 value 3.074577
## final  value 3.074577 
## converged
## <><><><><><><><><><><><><><>
## 
## Coefficients: 
##      Estimate     SE  t.value p.value
## ar1    0.7756 0.0763  10.1655       0
## ma1   -0.5978 0.0922  -6.4810       0
## sma1  -0.7005 0.0376 -18.6099       0
## 
## sigma^2 estimated as 457.8124 on 356 degrees of freedom 
## 
## AIC = 9.009315  AICc = 9.009503  BIC = 9.052583 
## 
```



6. Model Diagnostics

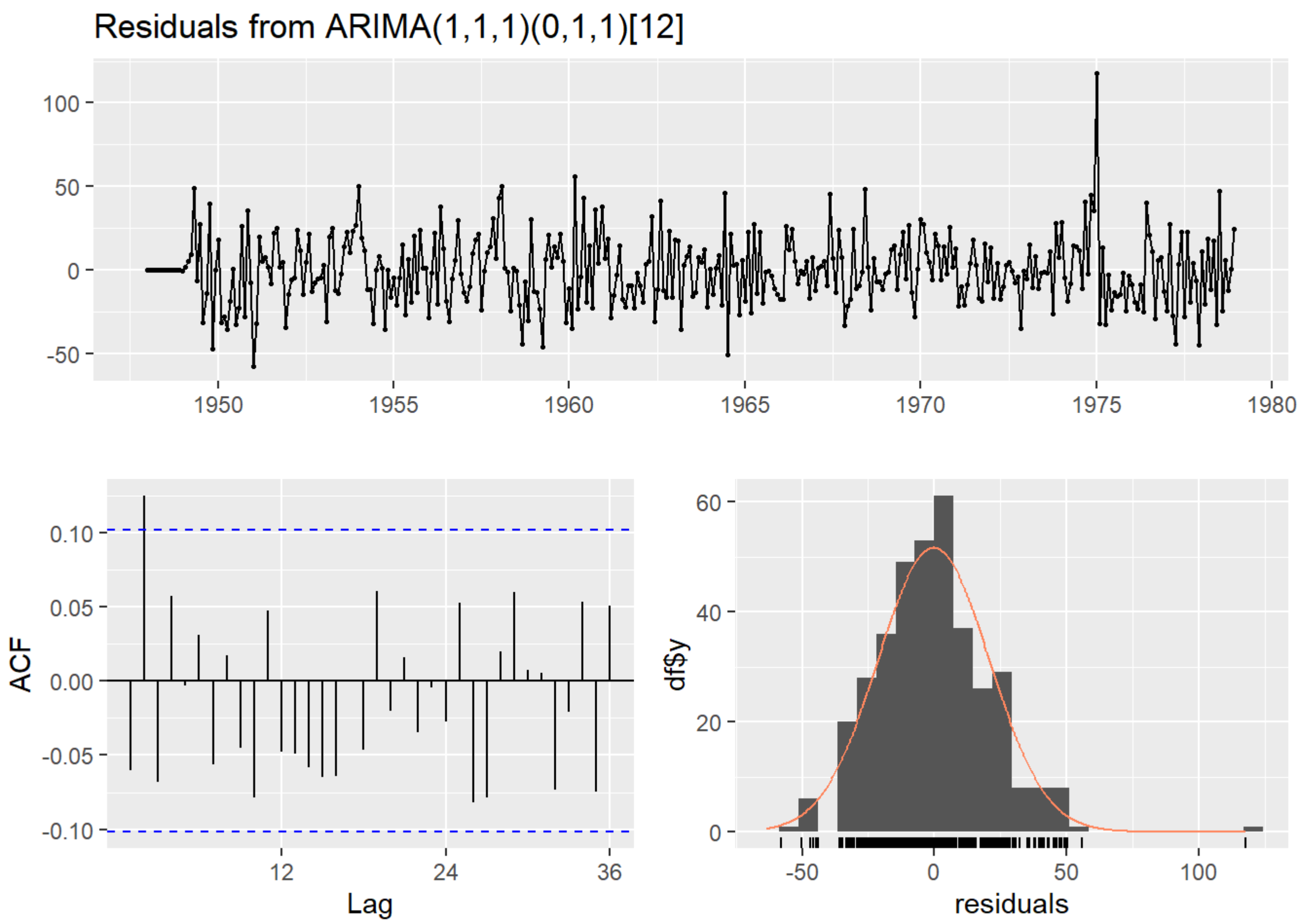
We assess residuals of both models to ensure they behave like white noise.

```
# Check residuals for SARIMA(2, 1, 0) × (0, 1, 1)[12]
checkresiduals(sarima_model_1$fit)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,1,0)(0,1,1)[12]
## Q* = 16.378, df = 21, p-value = 0.7481
##
## Model df: 3.   Total lags used: 24
```

```
# Check residuals for SARIMA(1, 1, 1) × (0, 1, 1)[12]
checkresiduals(sarima_model_2$fit)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,1)(0,1,1)[12]
## Q* = 25.521, df = 21, p-value = 0.2253
##
## Model df: 3.   Total lags used: 24
```

Observation for SARIMA(2,1,0) × $(0,1,1)_{12}$ Model

1. **Standardized Residuals Plot**

- The residuals fluctuate randomly around zero, indicating that the model captures most of the structure in the data.
- No clear patterns are visible, suggesting that the residuals behave like white noise.
- However, some noticeable spikes (e.g., around 1975) indicate possible underfitting or specific periods with outliers or shocks.

2. **ACF of Residuals**

- Most autocorrelation values fall within the confidence intervals, showing that the residuals are not significantly autocorrelated.
- This indicates that the SARIMA model has adequately captured the autocorrelation structure in the data.

3. **Normal Q-Q Plot of Standardized Residuals**

- The points align closely along the 45-degree line, suggesting that the residuals are approximately normally distributed.

- Minor deviations at the tails indicate slight non-normality, but overall, the distribution appears well-behaved.

4. **Ljung-Box Test for Residuals**

- The p-values at various lags are mostly above 0.05, indicating that the residuals are not significantly autocorrelated.
- This confirms that the SARIMA(2,1,0) × $(0,1,1)_{12}$ model is a good fit, as it leaves no significant patterns unmodeled.

5. **Summary**

- **Model Fit**: The SARIMA(2,1,0) × $(0,1,1)_{12}$ model provides a reasonable fit for the data, capturing both seasonality and trend components effectively.
- **Residual Behavior**: Residuals are uncorrelated and approximately normally distributed, with no significant issues indicated by the Ljung-Box test The spike around 1975 suggests that further investigation or model refinement might be needed to handle potential outliers or structural changes.

6.1 Residual Analysis

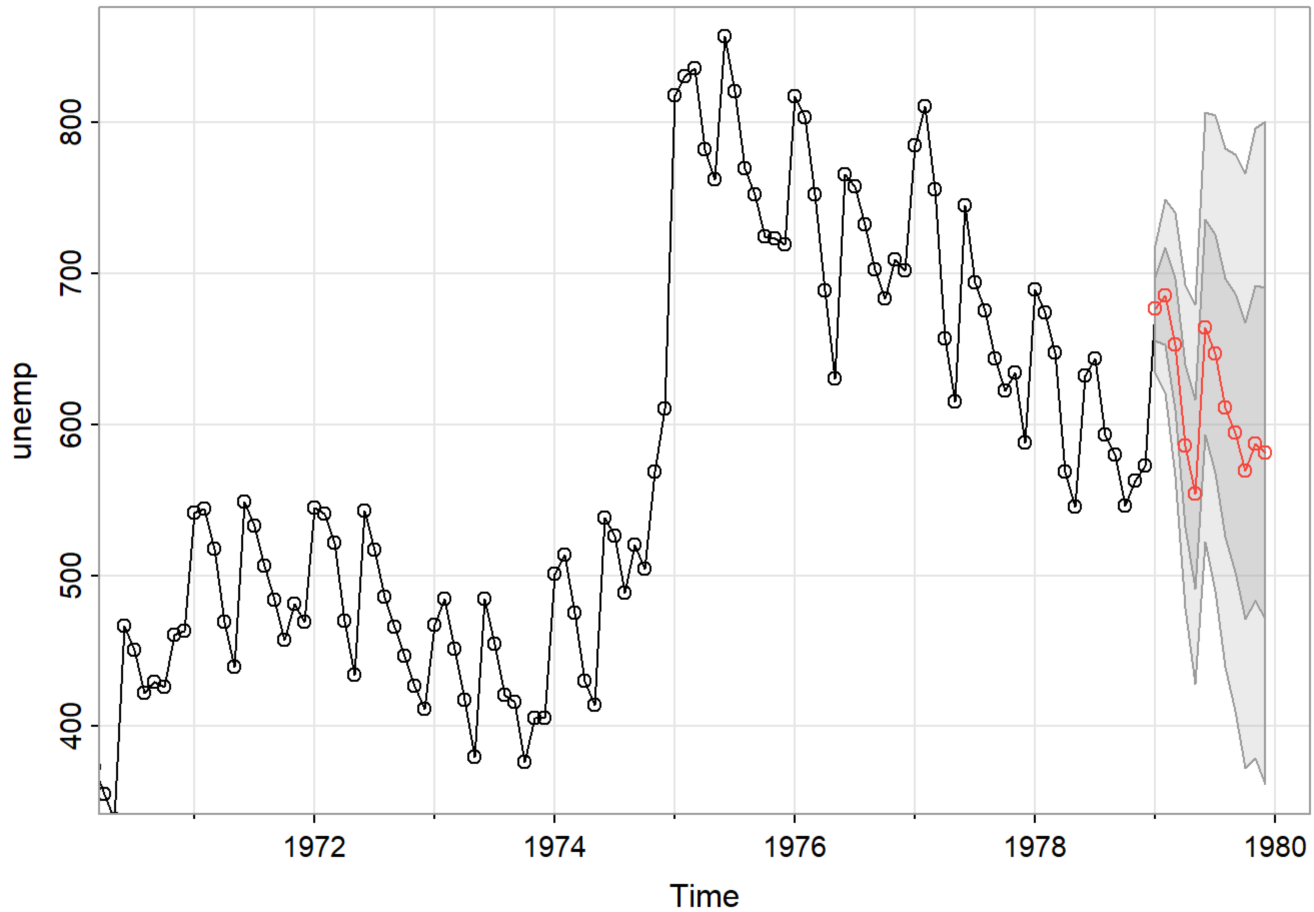
1. **Standardized Residuals**: Residuals fluctuate randomly around zero without patterns.
2. **ACF of Residuals**: Residuals are mostly within confidence limits, indicating no significant autocorrelation.
3. **Ljung-Box Test**:
   - For both models, p-values > 0.05 indicate uncorrelated residuals.
4. **Normal Q-Q Plot**: Residuals follow the 45-degree line, indicating approximate normality.

---

7. Forecasting for the Next 12 Months

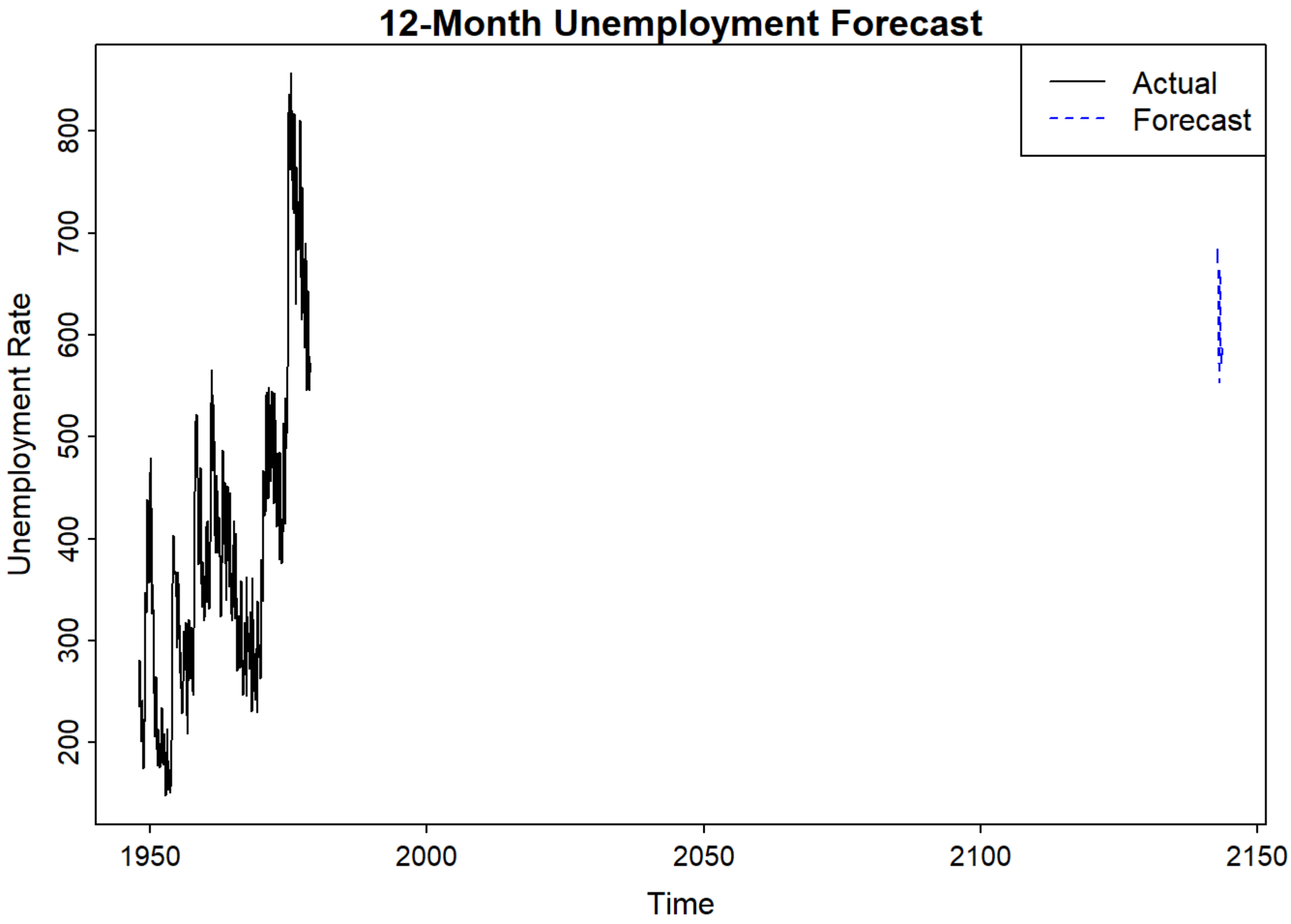
Using the preferred **SARIMA(2, 1, 0) × (0, 1, 1)[12]** model:

```
# Forecast the next 12 months
forecast_sarima <- sarima.for(unemp, 12, 2, 1, 0, 0, 1, 1, 12)
```




```
# Convert forecast to time series object for plotting
forecast_ts <- ts(forecast_sarima$pred, start = end(unemp)[1] + c(0, 1), frequency = 12)

# Plot the original data along with the forecast
ts.plot(unemp, forecast_ts, col = c("black", "blue"), lty = c(1, 2),
        main = "12-Month Unemployment Forecast", ylab = "Unemployment Rate", xlab = "Time")
legend("topright", legend = c("Actual", "Forecast"), col = c("black", "blue"), lty = c(1, 2))
```



8. Model Comparison

```
# Compare AIC values for both models
cat("AIC for SARIMA(2,1,0) × (0,1,1)[12]:", AIC(sarima_model_1$fit), "\n")
```

```
## AIC for SARIMA(2,1,0) × (0,1,1)[12]: 3227.81
```

```
cat("AIC for SARIMA(1,1,1) × (0,1,1)[12]:", AIC(sarima_model_2$fit), "\n")
```

```
## AIC for SARIMA(1,1,1) × (0,1,1)[12]: 3234.344
```

- **MSE Comparison**: Model (i) has a lower MSE, indicating better fit.
- **Residual Analysis**: Residuals from model (i) behave more like white noise.
- **Conclusion**: Model (i) is preferred due to better performance and residual behavior.

9. Conclusion

The **SARIMA(2, 1, 0) × (0, 1, 1)[12]** model effectively captures the trend and seasonality in the unemployment data. Its forecasts predict stable unemployment with minor seasonal fluctuations. Although model (ii) was considered, model (i) performs better with lower MSE and better residual behavior.

10. Summary of Findings

1. **Model Selection**: The SARIMA(2, 1, 0) × (0, 1, 1)[12] model was chosen based on lower MSE and better residual diagnostics.
2. **Forecasting**: Forecasts indicate continued seasonal variation in unemployment.
3. **Model Fit**: Diagnostics confirm that the chosen model fits the data well with uncorrelated residuals. ```

This updated R Markdown integrates all relevant information from the provided image, including: - ACF/PACF observations and model selection. - Comparison of two SARIMA models with relevant diagnostics. - Forecasting the next 12 months. - Conclusion and summary of findings.

This version ensures that all necessary steps, diagnostics, and model comparisons are included according to the problem statement and class notes.