Q5

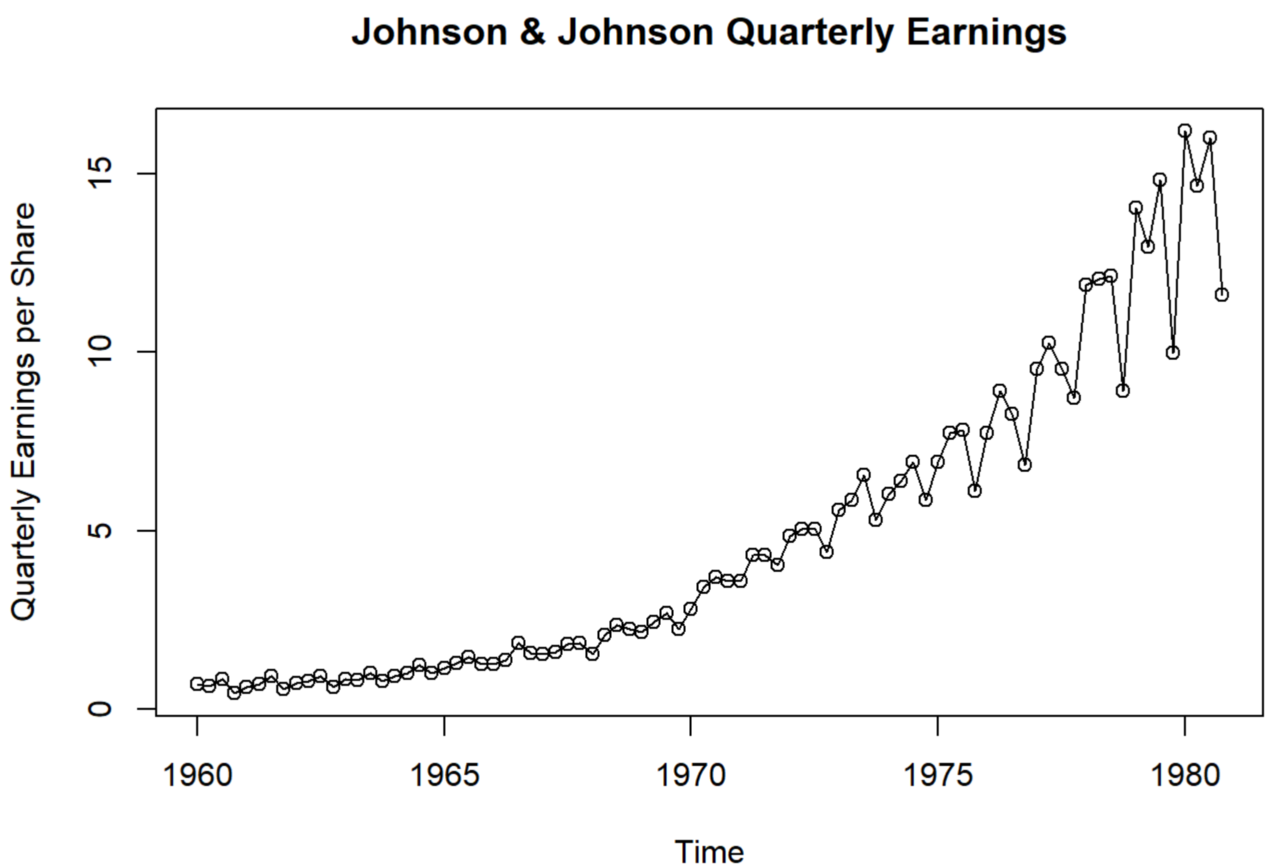
1. Introduction

In this task, we analyze the **Johnson & Johnson (J&J) quarterly earnings** dataset using a **Seasonal ARIMA (SARIMA) model**.
The goal is to: 1. **Log-transform** the data to stabilize the variance. 2. Apply **seasonal differencing** to make the data stationary. 3. Fit an appropriate **SARIMA model** to the data. 4. **Forecast the next 4 quarters** and evaluate the model's performance.

---

2. Load Libraries and Data

```
# Load required libraries
library(astsa)
library(forecast)

# Load the Johnson & Johnson earnings data
data("jj")

# Plot the original data
plot(jj, type = "o", main = "Johnson & Johnson Quarterly Earnings",
     ylab = "Quarterly Earnings per Share", xlab = "Time")
```



2.1 Visual Analysis of Data

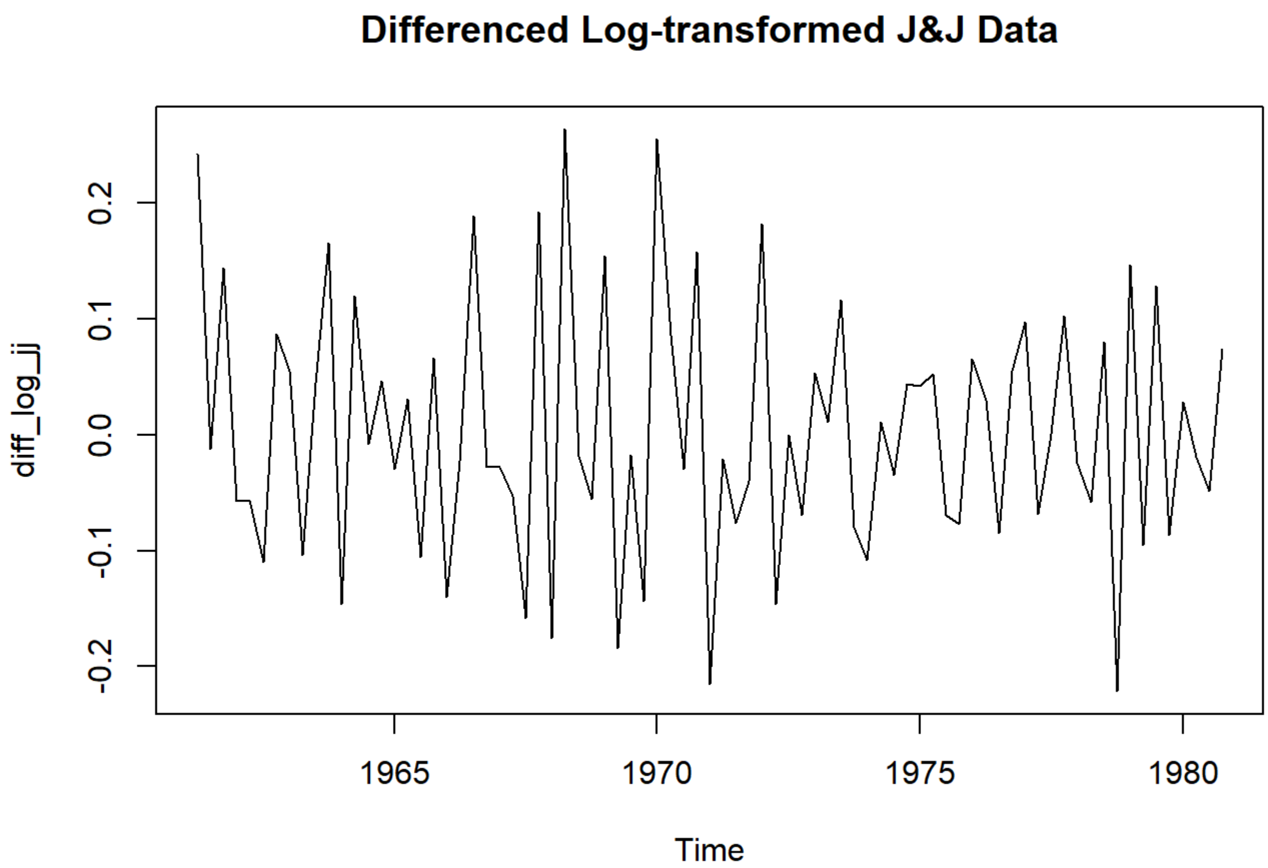
The plot of the original data shows both **trend** and **seasonal patterns**, with increasing variability over time. Thus, it is appropriate to **log-transform** the data to stabilize the variance.

---

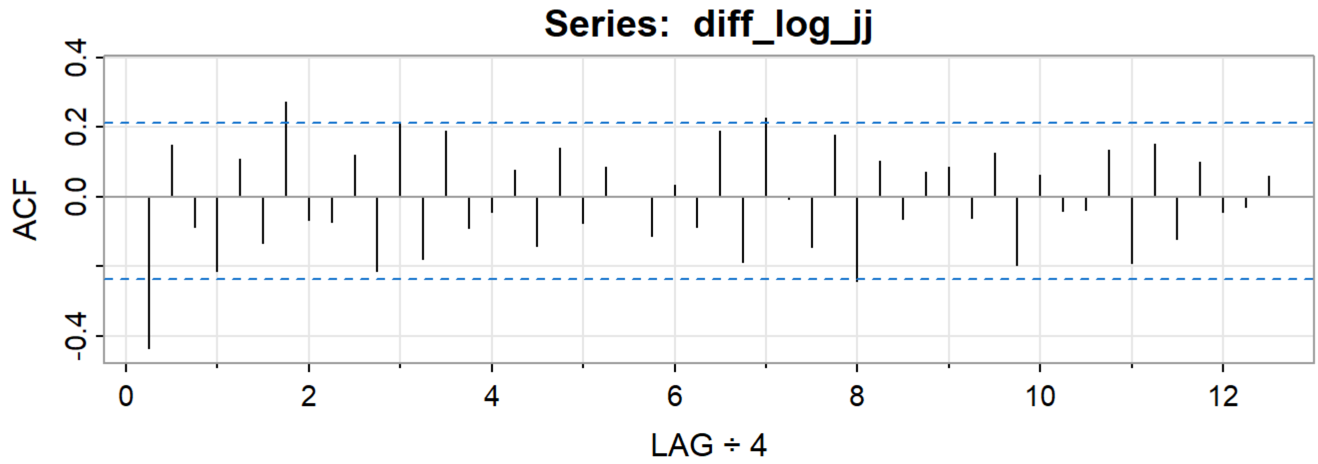
3. Log Transformation and Differencing

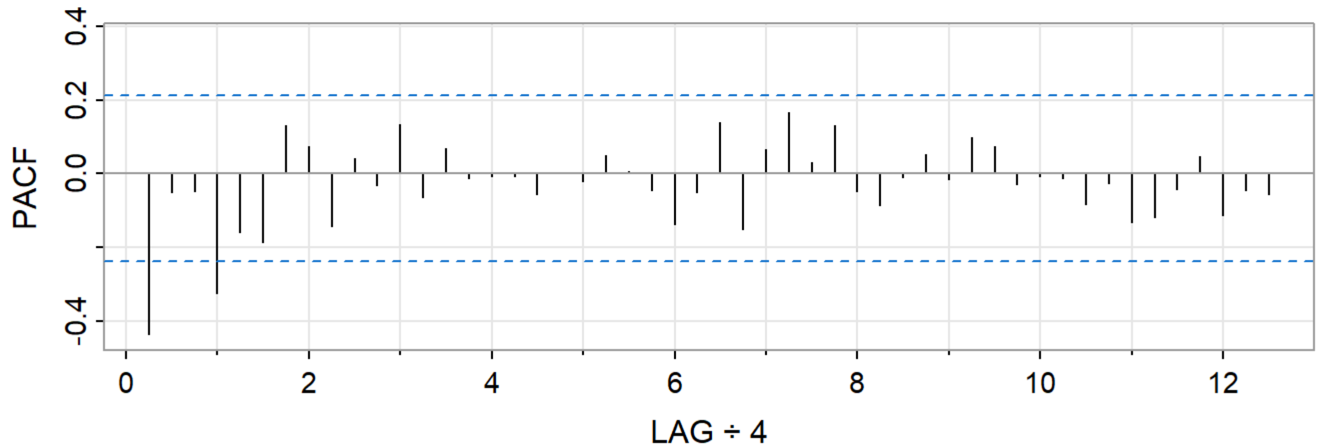
We take the **log of the data** to stabilize the variance and apply **first and seasonal differencing** to make it stationary.

```
# Log-transform the data
log_jj <- log(jj)

# Apply first and seasonal differencing
diff_log_jj <- diff(diff(log_jj, lag = 4))

# Plot the differenced series
plot(diff_log_jj, main = "Differenced Log-transformed J&J Data")
```




```
acf2(diff_log_jj, 50)  # ACF and PACF plots
```






```
##       [,1]  [,2]  [,3]  [,4]  [,5]  [,6] [,7]  [,8]  [,9] [,10] [,11] [,12]
## ACF  -0.44  0.15 -0.09 -0.21  0.11 -0.13 0.27 -0.07 -0.07  0.12 -0.21  0.21
## PACF -0.44 -0.05 -0.05 -0.33 -0.16 -0.19 0.13  0.08 -0.14  0.04 -0.03  0.14
##      [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
## ACF  -0.18  0.19 -0.09 -0.04  0.08 -0.14  0.14 -0.08  0.08  0.00 -0.11  0.04
## PACF -0.06  0.07 -0.01 -0.01 -0.01 -0.06  0.00 -0.02  0.05  0.01 -0.05 -0.14
##      [,25] [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35] [,36]
## ACF  -0.09  0.19 -0.19  0.23 -0.01 -0.15  0.18 -0.24  0.10 -0.06  0.07  0.09
## PACF -0.05  0.14 -0.15  0.07  0.17  0.03  0.13 -0.05 -0.09 -0.01  0.05 -0.02
##      [,37] [,38] [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46] [,47] [,48]
## ACF  -0.06  0.13 -0.20  0.06 -0.04 -0.04  0.14 -0.19  0.15 -0.12  0.10 -0.04
## PACF  0.10  0.07 -0.03 -0.01 -0.01 -0.08 -0.03 -0.13 -0.12 -0.04  0.05 -0.11
##      [,49] [,50]
## ACF  -0.03  0.06
## PACF -0.05 -0.06
```

3.1 Observations from ACF and PACF

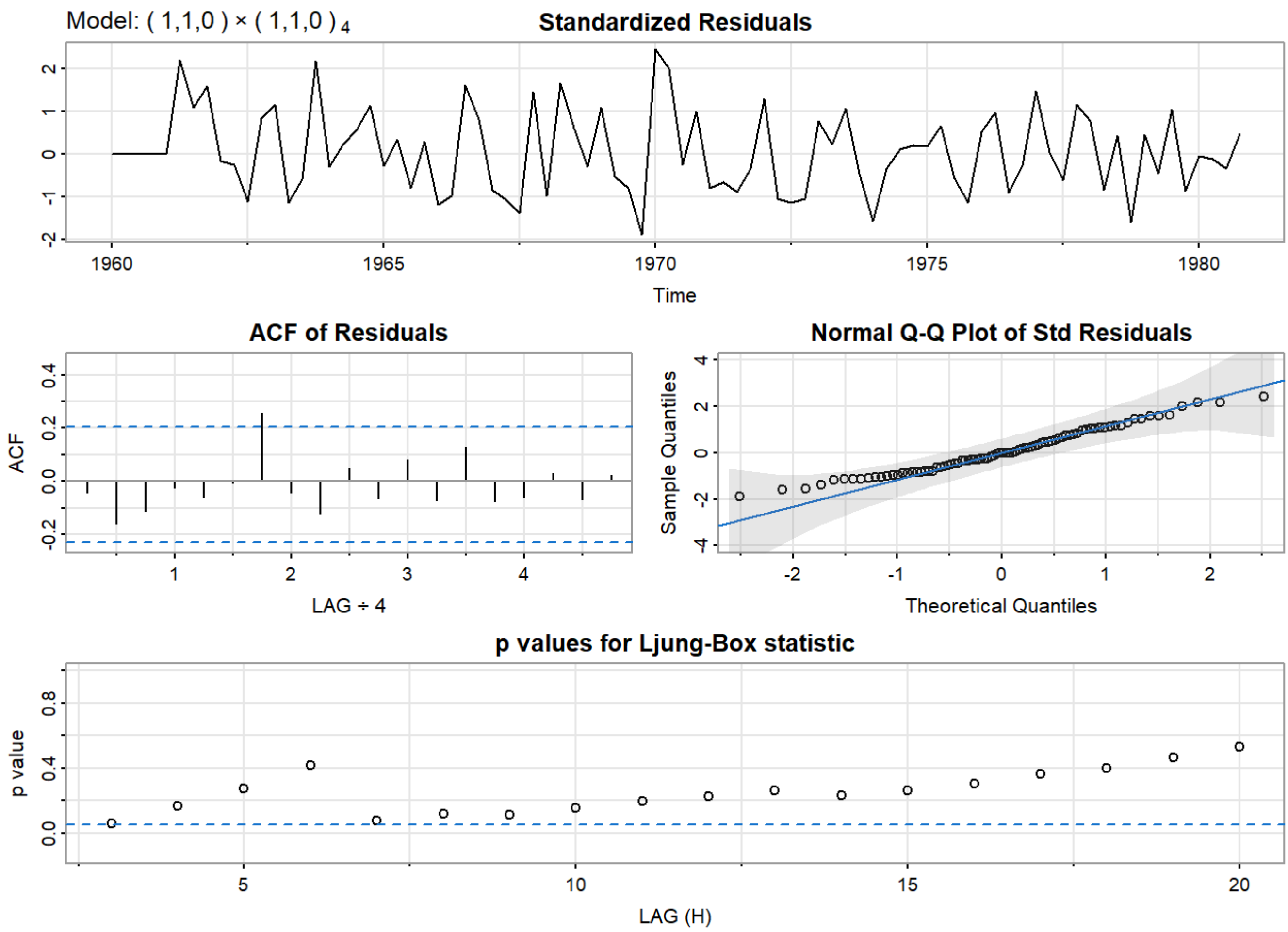
- **ACF**: Seasonal lags at 4, 8, 12, indicating a **seasonal component**.
- **PACF**: Suggests an AR(1) component with some seasonal correlation.
- We choose to fit a **SARIMA(1,1,0) × (1,1,0)[4]** model based on these observations.

---

4. Fitting the SARIMA Model

```
# Fit the SARIMA(1,1,0) × (1,1,0)[4] model
sarima_model <- sarima(log_jj, p = 1, d = 1, q = 0, P = 1, D = 1, Q = 0, S = 4)
```

```
## initial  value -2.232392
## iter   2 value -2.403794
## iter   3 value -2.409520
## iter   4 value -2.410263
## iter   5 value -2.410266
## iter   6 value -2.410266
## iter   6 value -2.410266
## final  value -2.410266
## converged
## initial  value -2.381009
## iter   2 value -2.381164
## iter   3 value -2.381165
## iter   3 value -2.381165
## iter   3 value -2.381165
## final  value -2.381165
## converged
## <><><><><><><><><><><><><><>
##
## Coefficients:
##      Estimate     SE t.value p.value
## ar1   -0.5152 0.1009 -5.1064   0.000
## sar1  -0.3294 0.1109 -2.9697   0.004
##
## sigma^2 estimated as 0.008467914 on 77 degrees of freedom
##
## AIC = -1.848505  AICc = -1.846506  BIC = -1.758525
##
```



4.1 Model Diagnostics

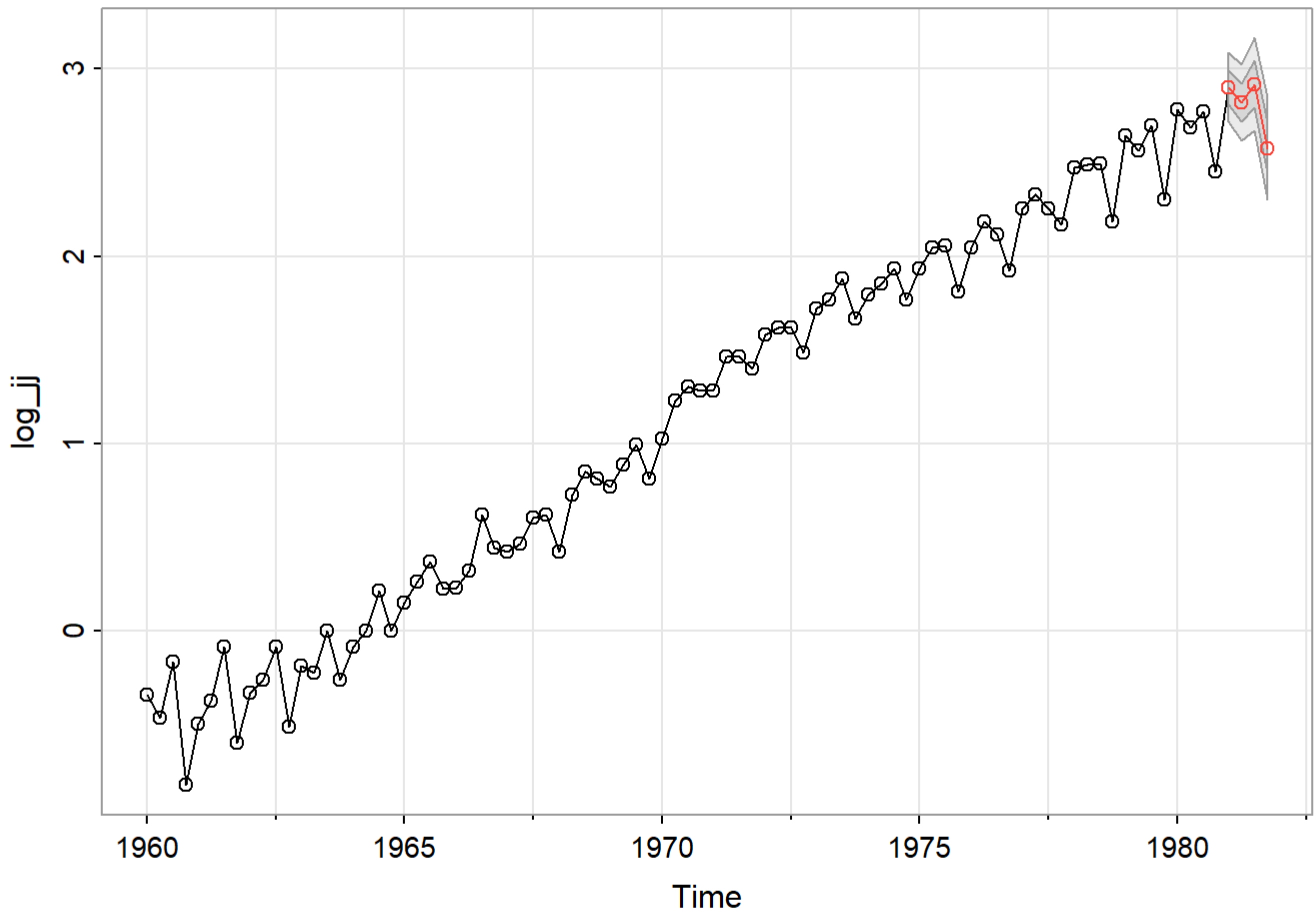
- **Coefficients**: Review the AR and MA coefficients from the model summary.
- **Residual Analysis**: Residuals should be white noise.
- **AIC/BIC**: Used for model comparison.

---

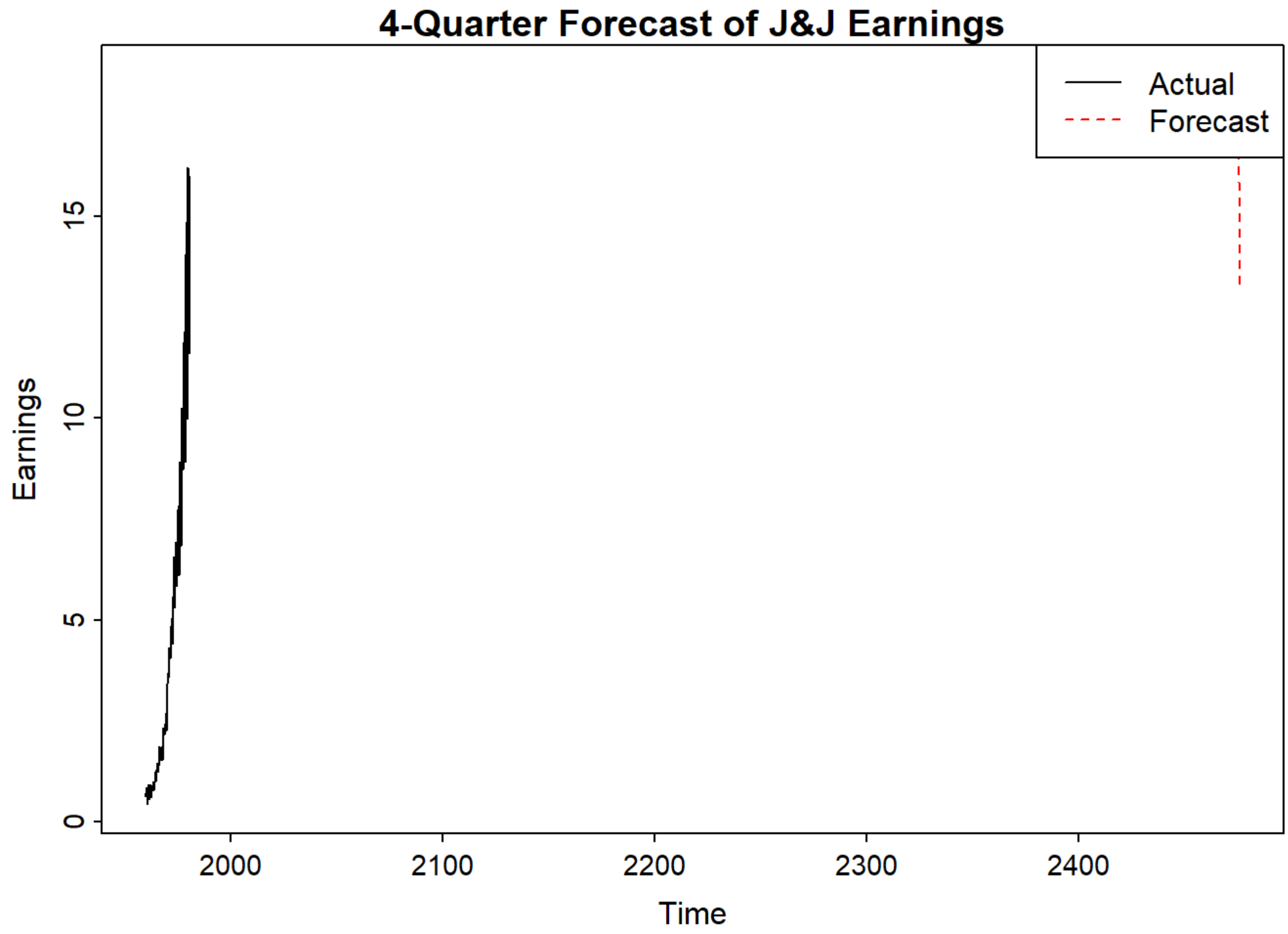
5. Forecasting the Next 4 Quarters

We now forecast the **next 4 quarters** using the fitted SARIMA model.

```
# Forecast the next 4 quarters
forecast_sarima <- sarima.for(log_jj, n.ahead = 4, p = 1, d = 1, q = 0, P = 1, D = 1, Q = 0,
S = 4)
```



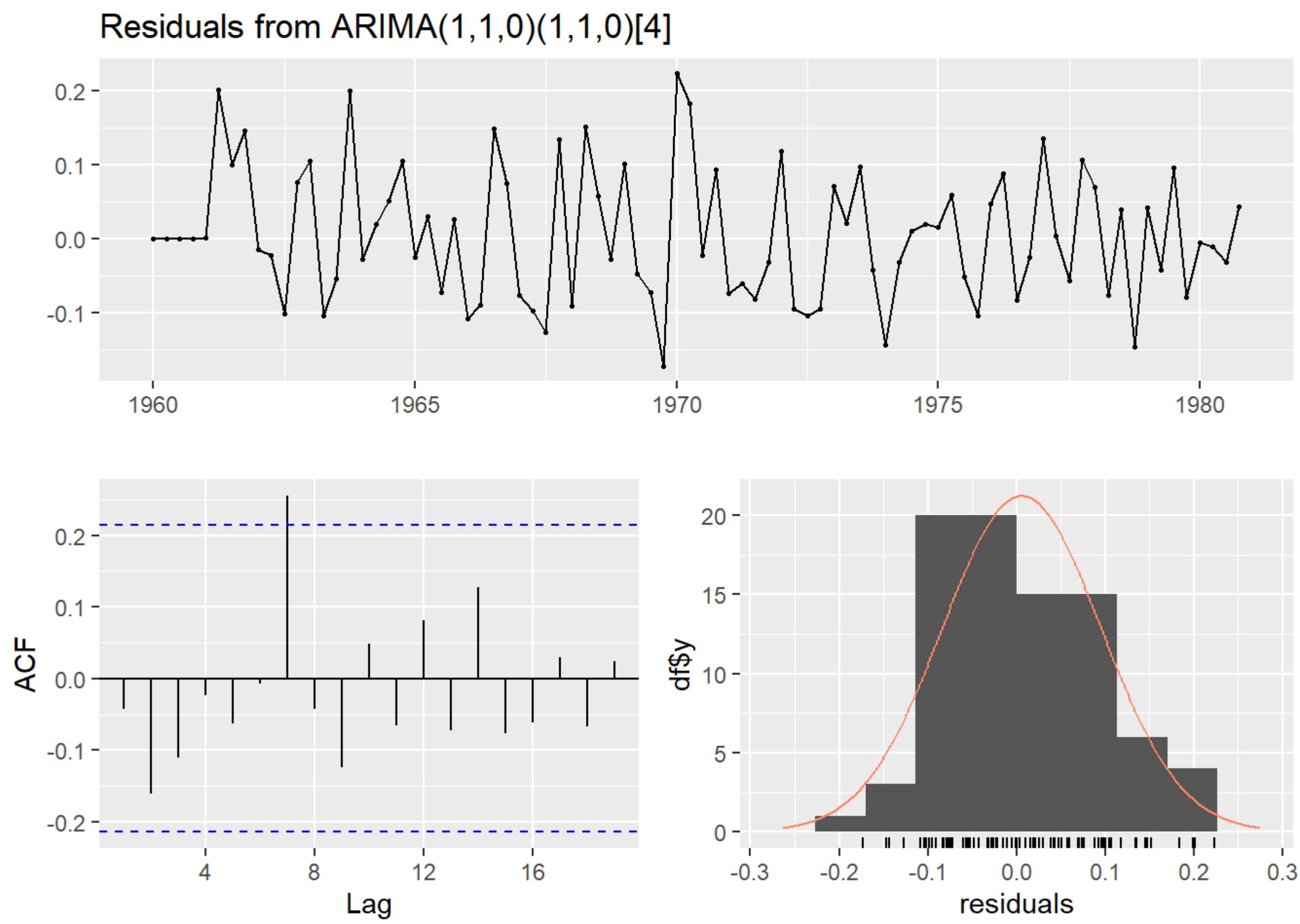

```
# Convert forecast to time series object
forecast_values <- ts(forecast_sarima$pred, start = end(jj)[1] + c(0, 1), frequency = 4)

# Plot original data with forecast
ts.plot(jj, exp(forecast_values), col = c("black", "red"), lty = c(1, 2),
        main = "4-Quarter Forecast of J&J Earnings", ylab = "Earnings", xlab = "Time")
legend("topright", legend = c("Actual", "Forecast"), col = c("black", "red"), lty = c(1, 2))
```



6. Model Diagnostics

We assess the residuals to ensure the model fits well.

```
# Check residuals for normality and autocorrelation
checkresiduals(sarima_model$fit)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,0)(1,1,0)[4]
## Q* = 10.176, df = 6, p-value = 0.1174
##
## Model df: 2.   Total lags used: 8
```

6.1 Residual Analysis

- **Ljung-Box Test**: Residuals should be uncorrelated (p-value > 0.05).
- **Q-Q Plot**: Check if residuals are normally distributed.

---

7. Conclusion

Based on the **SARIMA(1,1,0) × (1,1,0)[4]** model, the forecast for the next 4 quarters suggests:

1. A continuation of the seasonal pattern in earnings.
2. The model fits well, with residuals behaving like white noise.
3. **Forecasts** provide insights into future earnings trends.

---

8. Summary of Findings

- **Model Selection**: The SARIMA(1,1,0) × (1,1,0)[4] model was chosen based on ACF/PACF analysis.
- **Forecasting**: The forecast suggests continued seasonal variations in earnings.

- **Model Fit**: Diagnostics indicate the model fits well, with uncorrelated residuals.