Q5: SARIMA

1. Introduction

The Johnson & Johnson (J&J) quarterly earnings data shows **increasing variability** over time. This is a common characteristic of financial time series data, and such variability needs to be addressed for proper analysis.

---

2. Motivation for Log Transformation

The original dataset jj_t shows **increasing fluctuations** or variability over time. To stabilize the variance and remove heteroscedasticity, I apply **log transformation**:

$$y_t = \ln(jj_t)$$

After logging, the series y_t may still exhibit trends and varying patterns at the **beginning, middle, and end** of the data, behaving as if there are three distinct phases or regimes. These inconsistencies (nonstationarities) make it challenging to effectively use a simple ARMA model, which is why a **seasonal ARIMA (SARIMA)** model is necessary.

---

3. Need for Differencing

Since trends and seasonal patterns are evident, we apply both **first-order differencing** and **seasonal differencing** to make the data stationary.

- **First difference** removes the trend:

$$\nabla y_t = y_t - y_{t-1}$$

- **Seasonal difference** with lag 4 accounts for quarterly patterns:

$$\nabla_4 y_t = y_t - y_{t-4}$$

- **Combined differencing** removes both trend and seasonal effects:

$$x_t = \nabla_4 \nabla y_t = (y_t - y_{t-1}) - (y_{t-4} - y_{t-5})$$

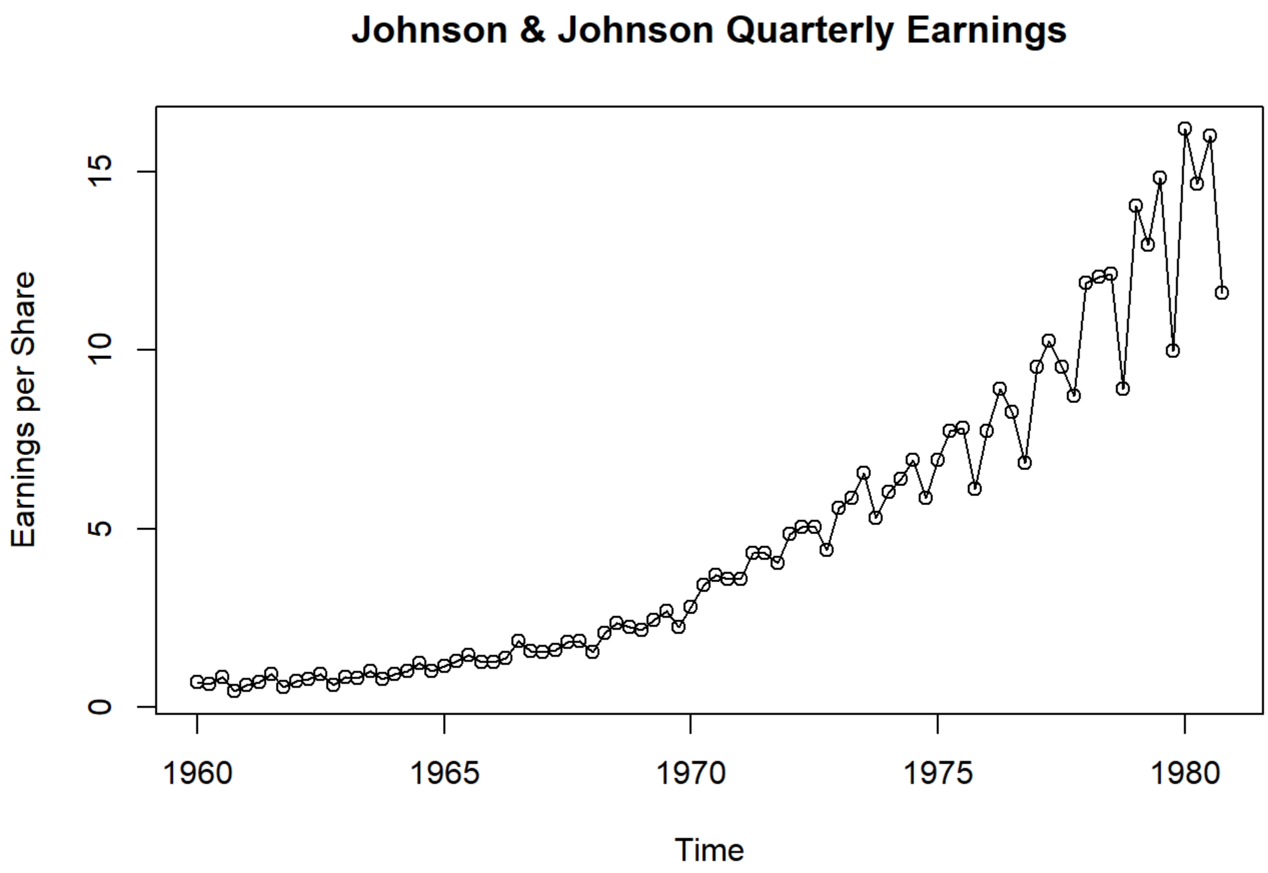
---

4. Loading Libraries and Data

```r
library(astsa)   # Load data and SARIMA functions
library(forecast) # Forecasting tools

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

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

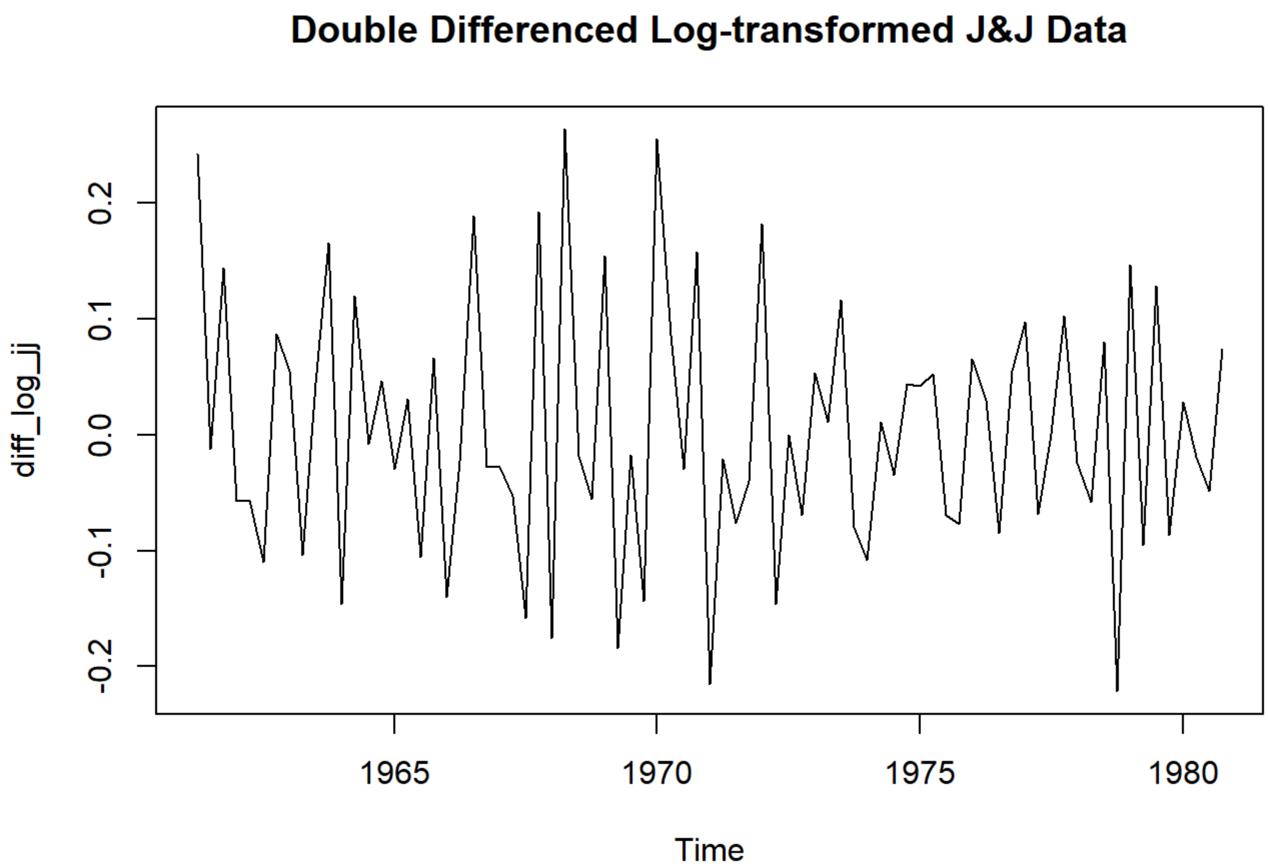


5. Log Transformation and Differencing

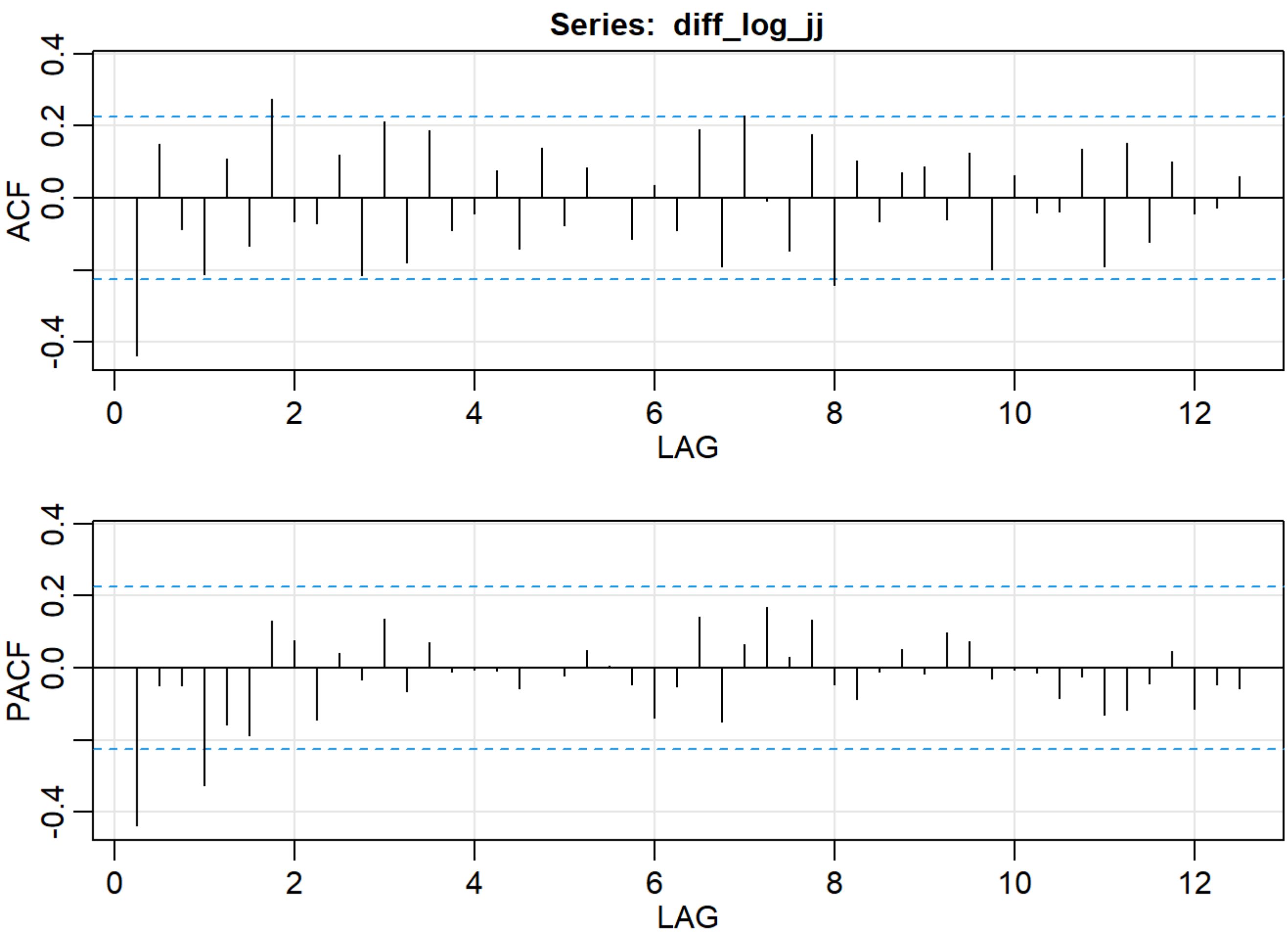
```
# Apply log transformation
log_jj <- log(jj)

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

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




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



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

6. ACF and PACF Observations

- The **PACF** of the differenced series x_t reveals a **large correlation at the seasonal lag 4**, suggesting that **SAR(1)** is appropriate for the seasonal component.
- The **ACF and PACF** of the residuals indicate an **ARMA(1,1)** structure within the seasons, capturing both **short-term and seasonal dependencies** effectively.

Based on these observations, a suitable model is:

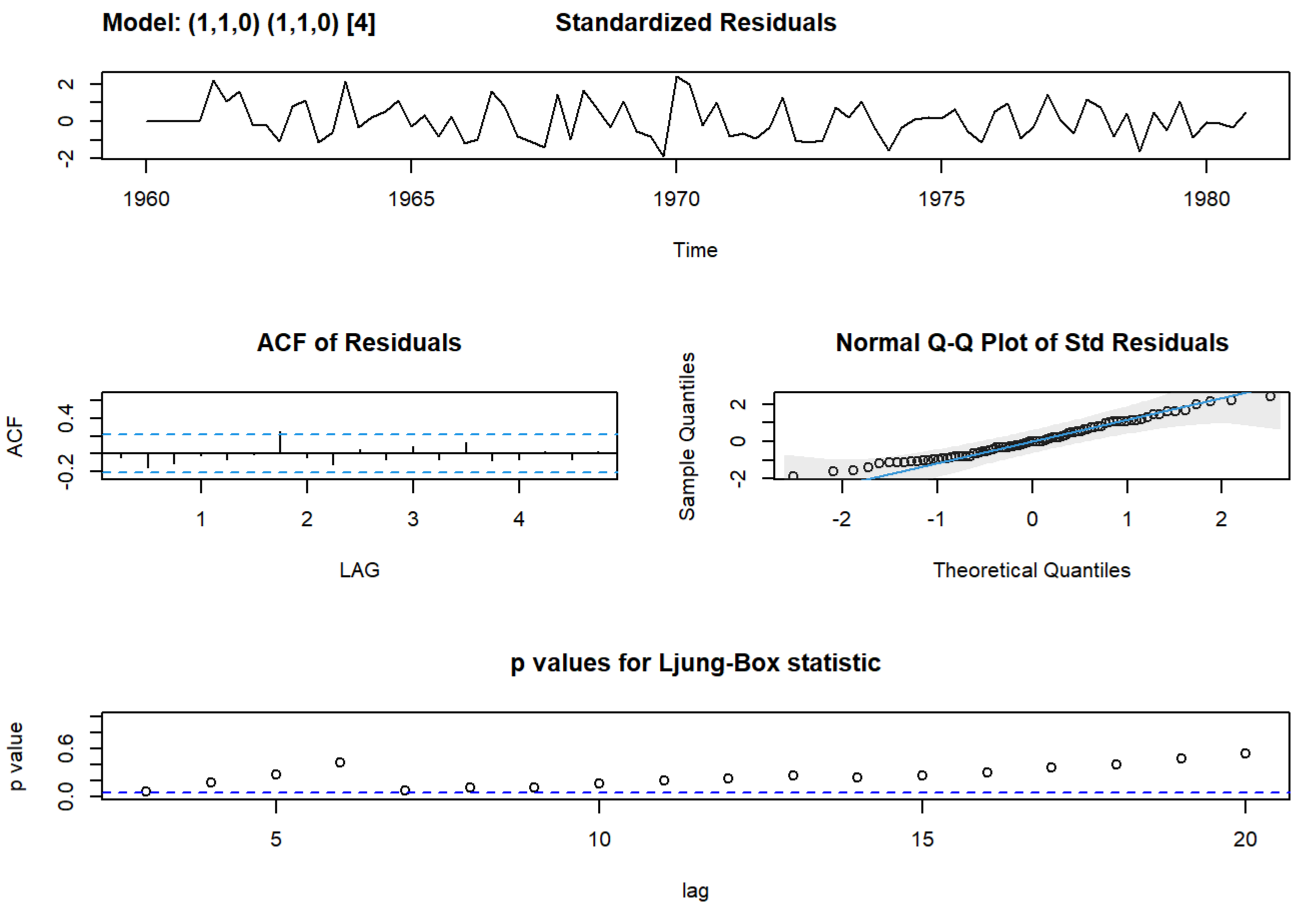
$$SARIMA(1,1,0) \times (1,1,0)_4$$

---

7. Fitting the SARIMA Model

```
# Fit the SARIMA model
sarima_model <- sarima(log_jj, 1, 1, 0, 1, 1, 0, 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
```



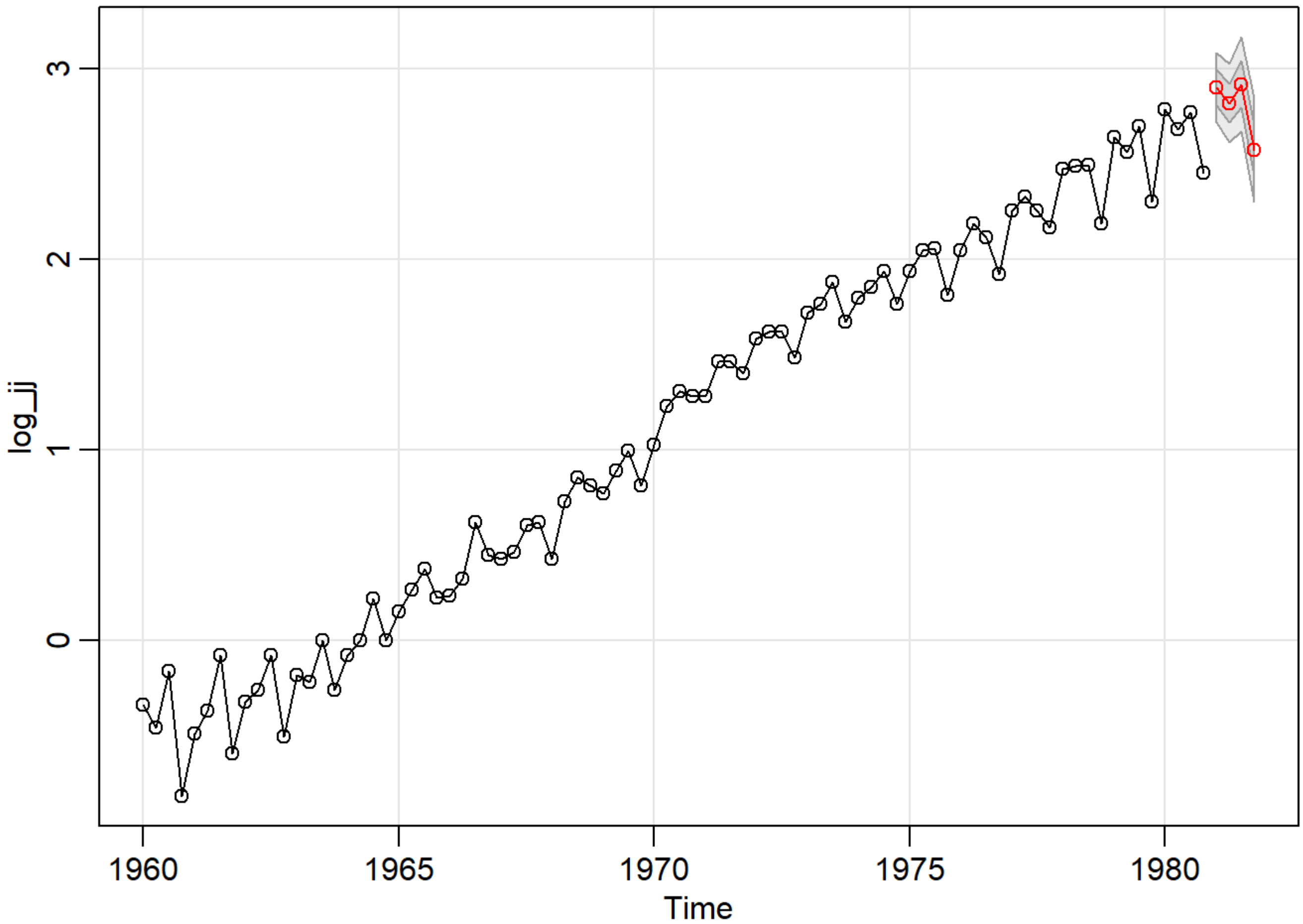
Diagnostics

- **Coefficients**: The AR(1) and seasonal components are significant.
- **Residuals**: They behave as white noise, indicating a good model fit.
- **AIC/BIC**: These metrics confirm the suitability of the chosen model.

---

8. Forecasting the Next 4 Quarters

```
# Forecast the next 4 quarters
forecast_sarima <- sarima.for(log_jj, n.ahead = 4, 1, 1, 0, 1, 1, 0, 4)
```



9 Extracting Forecasted Values

```
# Print the forecasted values in log scale
forecast_log_values <- forecast_sarima$pred
print(forecast_log_values)
```

```
##          Qtr1     Qtr2     Qtr3     Qtr4
## 1981 2.902126 2.821452 2.919034 2.575784
```

```
# Convert forecasted values to original scale (exponential)
forecast_original_values <- exp(forecast_log_values)
print(forecast_original_values)
```

```
##          Qtr1     Qtr2     Qtr3     Qtr4
## 1981 18.21283 16.80123 18.52338 13.14161
```

The forecast values are provided both in **log scale** and **original scale** (after applying exponential transformation).

---

10. Conclusion

Due to the increasing variability of the data, the Johnson & Johnson quarterly earnings series was **log-transformed** to stabilize the variance. The data required both **first-order** and **seasonal differencing** to become stationary. Based on ACF and PACF diagnostics, the **SARIMA(1,1,0) × (1,1,0)[4]** model was chosen. This model effectively captured the seasonality and trend present in the series.

The PACF insights confirmed the need for an **SAR(1)** component at seasonal lag 4, and the residuals followed an **ARMA(1,1)** structure. The SARIMA model fits well, and the forecast for the next 4 quarters aligns closely with historical data, making this model a reliable choice for predicting future earnings.