SARIMA fitting: Johnson & Johnson

PRACTICAL TIME SERIES ANALYSIS

THISTLETON AND SADIGOV

Objectives

- Fit SARIMA models to quarterly earnings of Johnson & Johnson share
- Forecast future values of examined time series

Modeling

- Time plot
- Transformation
- Differencing (seasonal or non-seasonal)
- Ljung-Box test
- ACF → Adjacent spikes → MA order
- ACF → Spikes around seasonal lags → SMA order
- PACF → Adjacent spikes → AR order
- PACF → Spikes around seasonal lags → SAR order

Modeling cont.

- Fit few different models
- Compare AIC, choose a model with minimum AIC
- **The parsimony principle**
- Time plot, ACF and PACF of residuals
- Ljung-Box test for residuals

The parsimony principle

SARIMA$(p, d, q, P, D, Q)_S$

$$p + d + q + P + D + Q \leq 6$$

JohnsonJohnson {datasets}- AGAIN

- Quarterly earnings (dollars) per Johnson & Johnson share 1960–80.
- Quarterly time series
- Source: “astsa” package

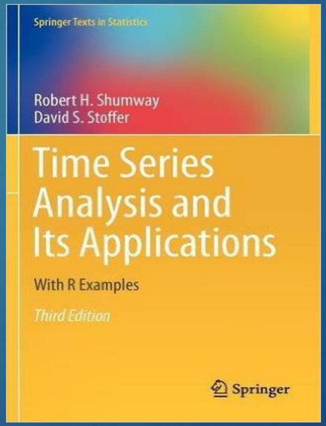
Shumway, R.H. and Stoffer, D.S. (2000)

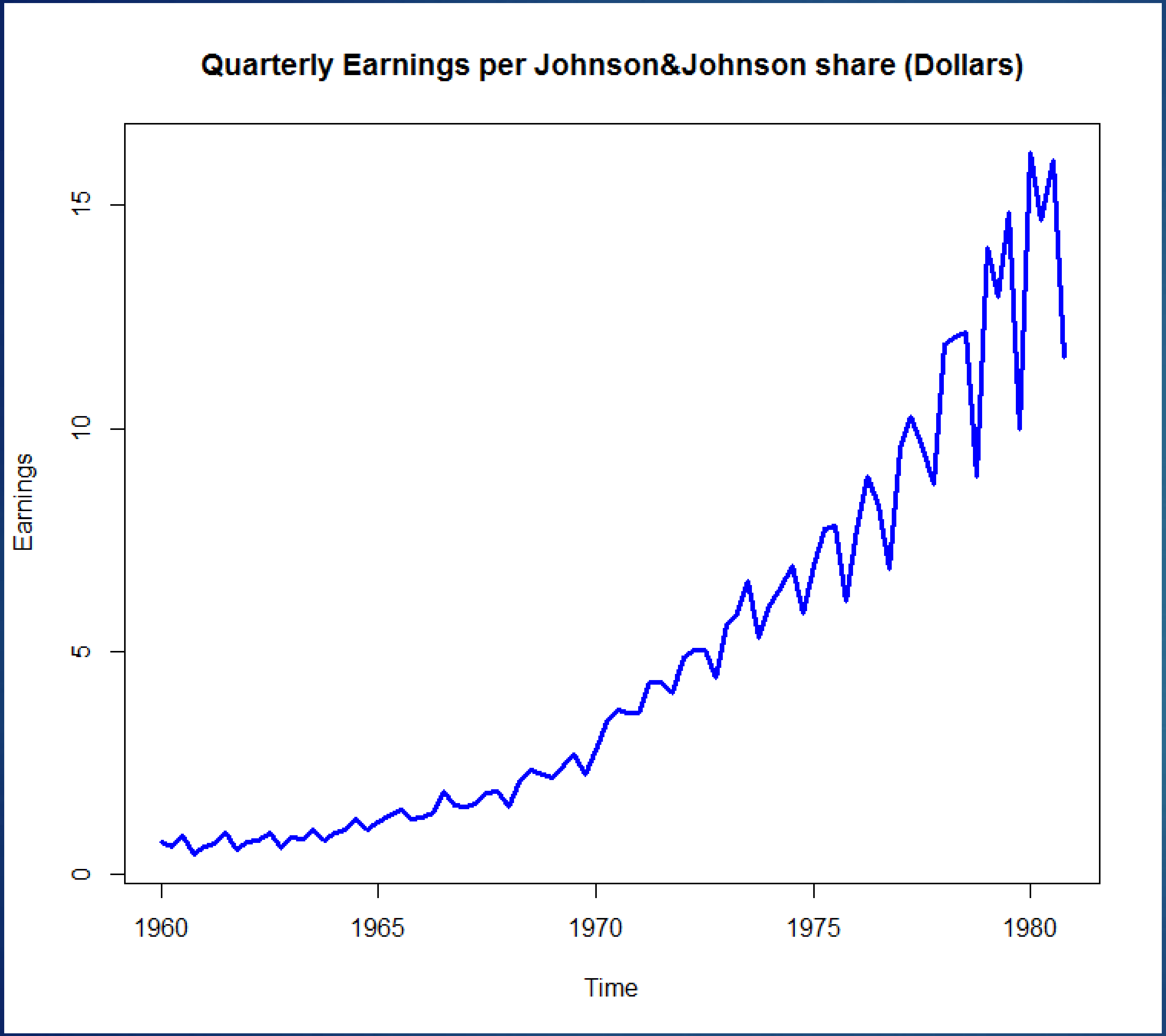
Time Series Analysis and its Applications

With R examples

Third Edition

Springer





Quarterly Earnings per Johnson&Johnson share (Dollars)
Earnings
0
5
10
15
1960
1965
1970
1975
1980
Time

Transformation

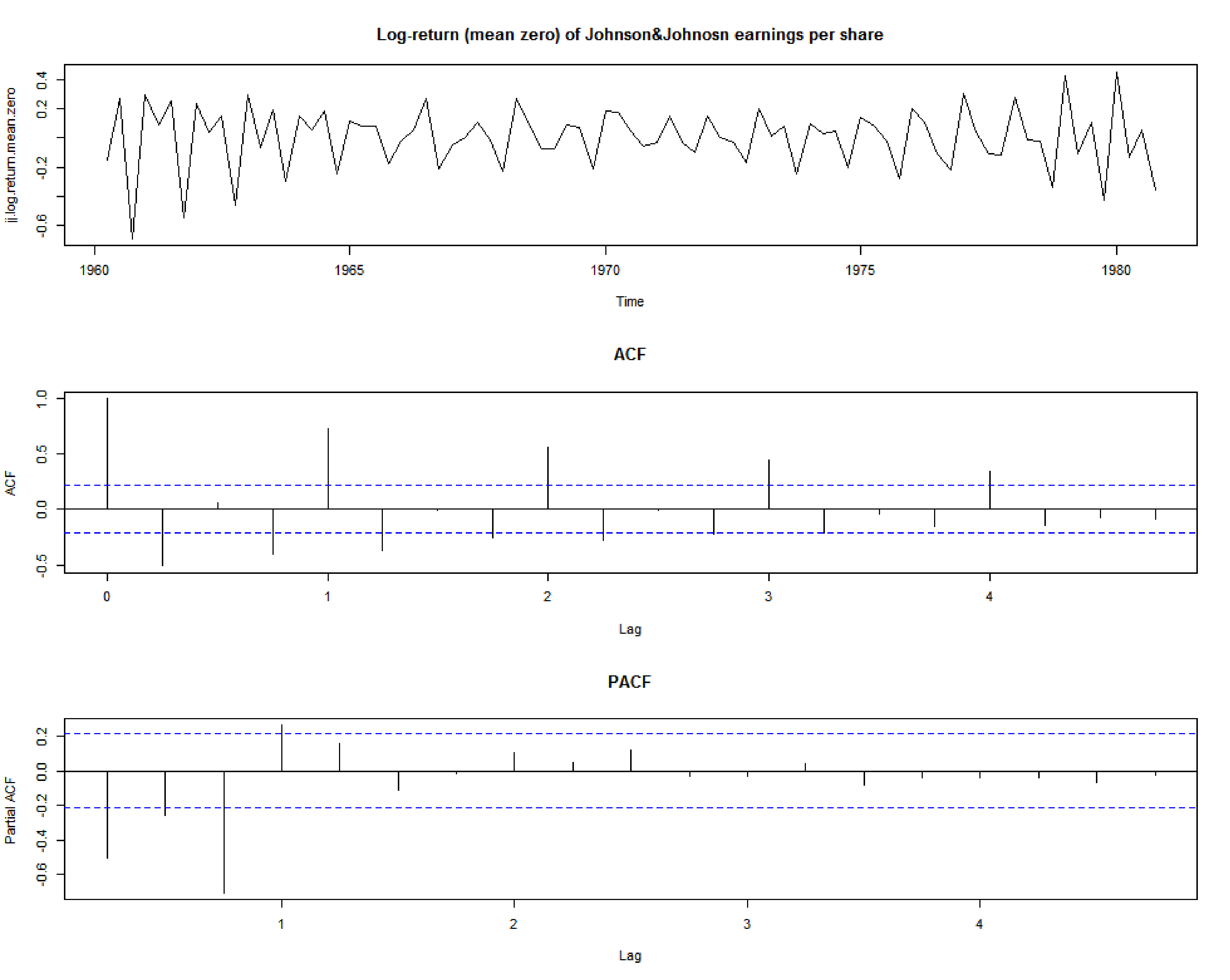
Log-return a time series $\{X_t\}$

is defined as

$$r_t = \log\left(\frac{X_t}{X_{t-1}}\right) = \log(X_t) - \log(X_{t-1})$$

In R,

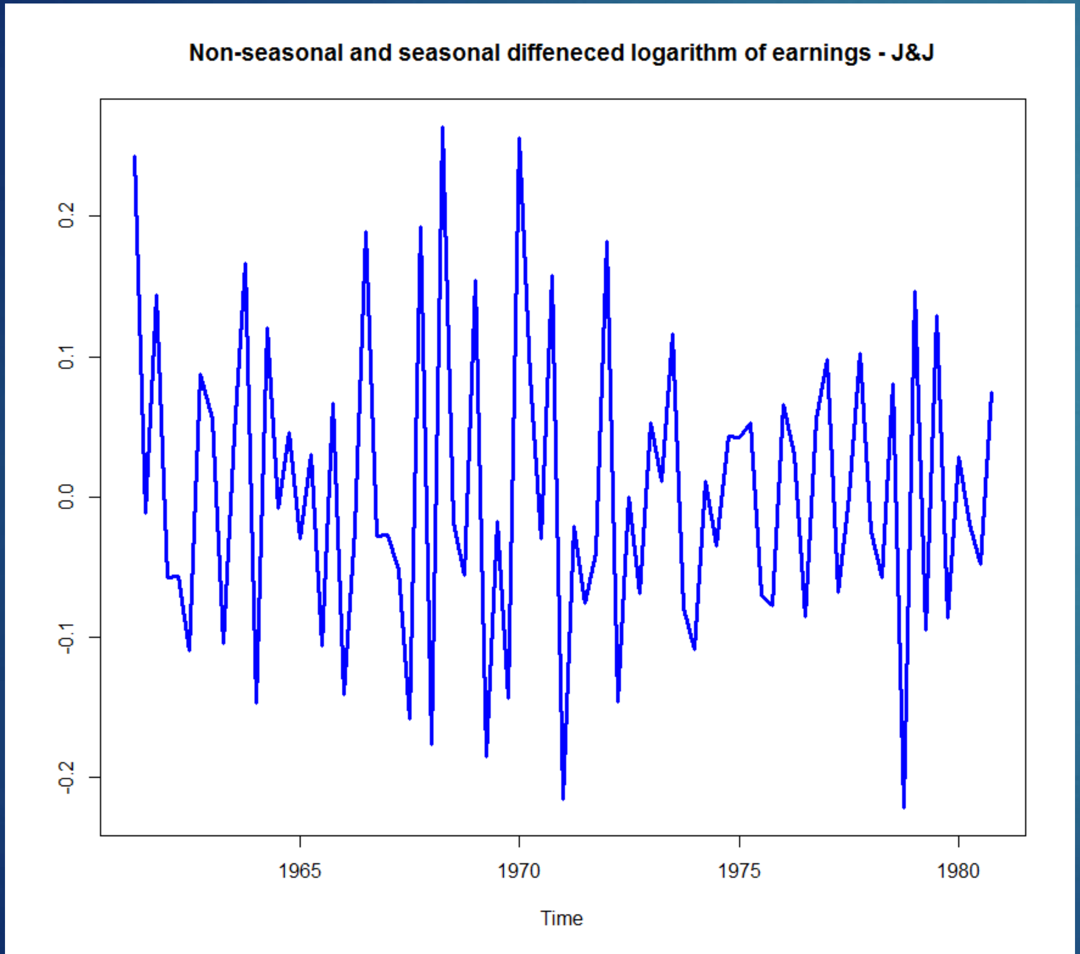
$$diff(\log(\quad))$$



Log-return (mean zero) of Johnson&Johnosn earnings per share
jj.log.return.mean.zero
0.4
0.2
-0.2
-0.6
1960
1965
1970
1975
1980
Time
ACF
1.0
0.5
0.0
-0.5
ACF
0
1
2
3
4
Lag
PACF
0.2
0.0
-0.2
-0.4
-0.6
Partial ACF
1
2
3
4
Lag

Seasonal differencing D=1

$$diff(diff(\log(jj)), 4)$$



Ljung-Box test

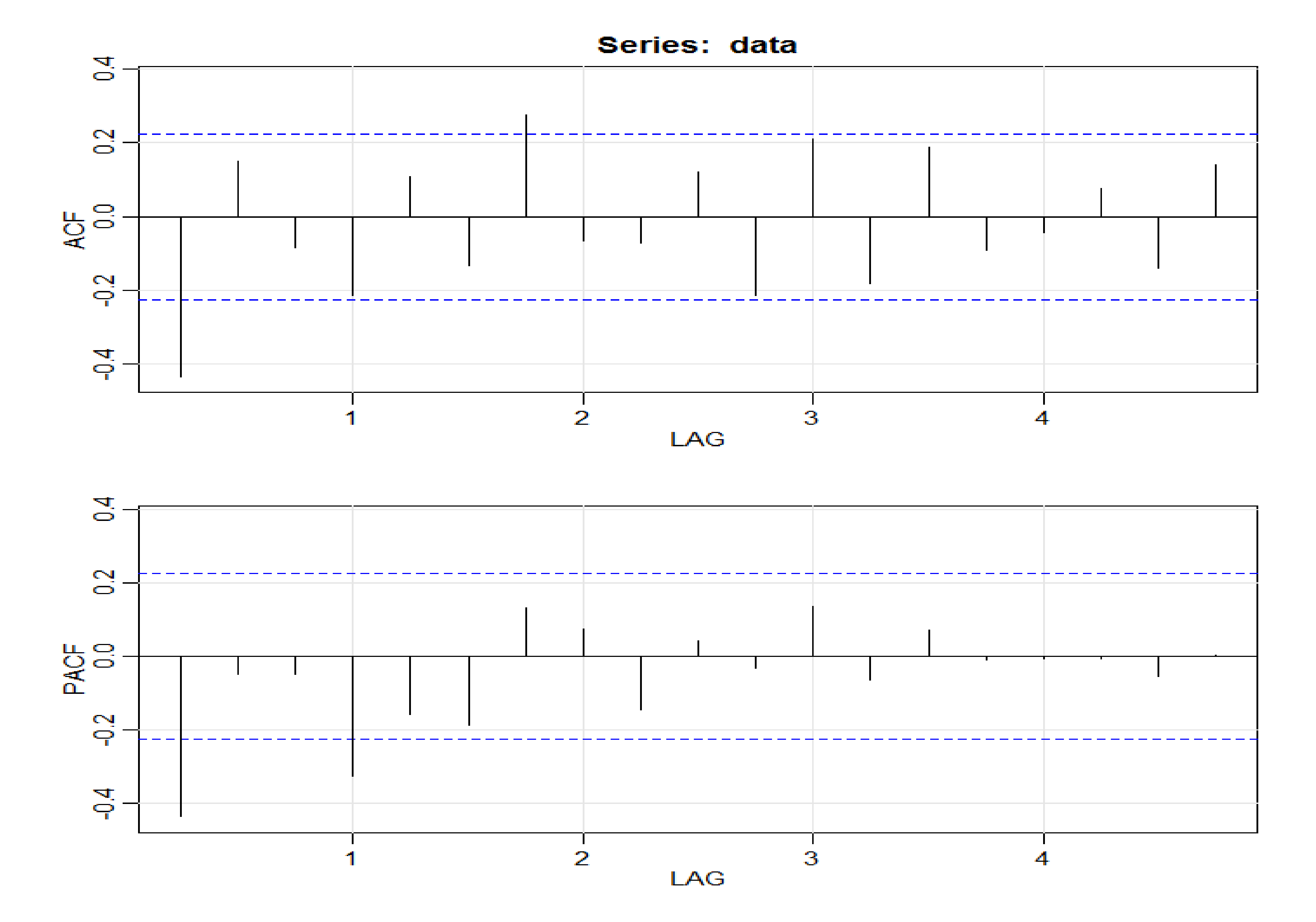
- Box.test(data, lag=log(length(data)))
- p-value:

$$0.0004658$$

So, we reject the hypothesis that there is no autocorrelation between previous lags of seasonal and non-seasonal differenced logarithm of earnings per J&J share

ACF

PACF


Series: data
ACF
0.4
0.2
0.0
-0.2
-0.4
1
2
3
4
LAG
PACF
0.4
0.2
0.0
-0.2
-0.4
1
2
3
4
LAG

Order specification and parameter estimation

- ACF → $q = 0{,}1\ ;\ Q = 0, 1$
- PACF → $p = 0, 1;\ \mathrm{P} = 0\ , 1$
- So, we will look at SARIMA$(p, 1, q, P, 1, Q)_4$ modeLS for $\log(jj)$ where

$$0 \leq p, q, P, Q \leq 1$$

- R routine:

$$arima(x = \log(jj), order\ =\ c(p, 1, q), seasonal\ =\ list(order = c(P, 1, Q), period = 4))$$

0 1 0 0 1 0 4 AIC= -124.0685 SSE= 0.9377871 p-VALUE= 0.0002610795

0 1 0 0 1 1 4 AIC= -126.3493 SSE= 0.8856994 p-VALUE= 0.0001606542

0 1 0 1 1 0 4 AIC= -125.9198 SSE= 0.8908544 p-VALUE= 0.0001978052

0 1 0 1 1 1 4 AIC= -124.3648 SSE= 0.8854554 p-VALUE= 0.000157403

0 1 1 0 1 0 4 AIC= -145.5139 SSE= 0.6891988 p-VALUE= 0.03543717

0 1 1 0 1 1 4 AIC= -150.7528 SSE= 0.6265214 p-VALUE= 0.6089542

0 1 1 1 1 0 4 AIC= -150.9134 SSE= 0.6251634 p-VALUE= 0.7079173

0 1 1 1 1 1 4 AIC= -149.1317 SSE= 0.6232876 p-VALUE= 0.6780876

1 1 0 0 1 0 4 AIC= -139.8248 SSE= 0.7467494 p-VALUE= 0.03503386

1 1 0 0 1 1 4 AIC= -146.0191 SSE= 0.6692691 p-VALUE= 0.5400205

1 1 0 1 1 0 4 AIC= -146.0319 SSE= 0.6689661 p-VALUE= 0.5612964

1 1 0 1 1 1 4 AIC= -144.3766 SSE= 0.6658382 p-VALUE= 0.5459445

1 1 1 0 1 0 4 AIC= -145.8284 SSE= 0.667109 p-VALUE= 0.2200484

1 1 1 0 1 1 4 AIC= -148.7706 SSE= 0.6263677 p-VALUE= 0.594822

1 1 1 1 1 0 4 AIC= -148.9175 SSE= 0.6251104 p-VALUE= 0.7195469

1 1 1 1 1 1 4 AIC= -144.4483 **SSE= 0.6097742** p-VALUE= 0.3002702

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

Fit this model

$$X_t = Earnings$$

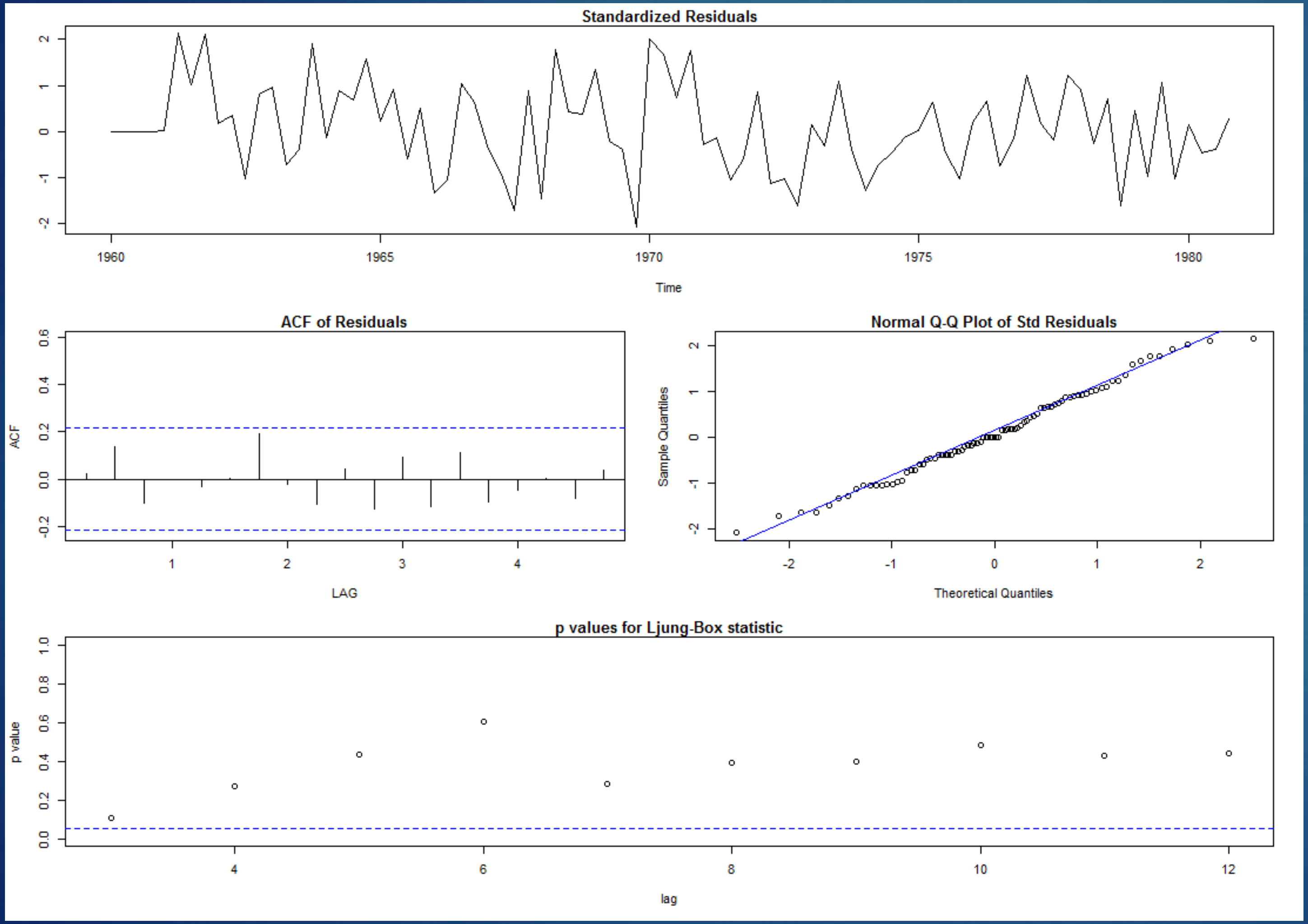
$$Y_t = \log(X_t)$$

	Estimate	SE	t.value	p.value
ma1	-0.6796	0.0969	-7.0104	0.0000
sar1	-0.3220	0.1124	-2.8641	0.0053

SARIMA routine

- 'astsa' package
- sarima$(\log(jj), 0{,}1{,}1{,}1{,}1{,}0{,}4)$

Residual analysis



Model – SARIMA$(0,1,1,1,1,0)_4$

$$X_t = Earnings$$

$$Y_t = \log(X_t)$$

$$(1-B)(1-B^4)(1-\Phi B^4)Y_t = (1+\theta B)Z_t$$

$$Y_t = Y_{t-1} + (\Phi+1)Y_{t-4} - (\Phi+1)Y_{t-5} - \Phi Y_{t-8} + \Phi Y_{t-9} + Z_t + \theta Z_{t-1}$$

	Estimate	SE	t.value	p.value
ma1	-0.6796	0.0969	-7.0104	0.0000
sar1	-0.3220	0.1124	-2.8641	0.0053

Model – cont.

$$Y_t = Y_{t-1} + 0.6780\ Y_{t-4} - 0.6780\ Y_{t-5} + 0.3220\ Y_{t-8} - 0.3220\ Y_{t-9} + Z_t - 0.6796\ Z_{t-1}$$

where

$$Y_t = \log(X_t)$$

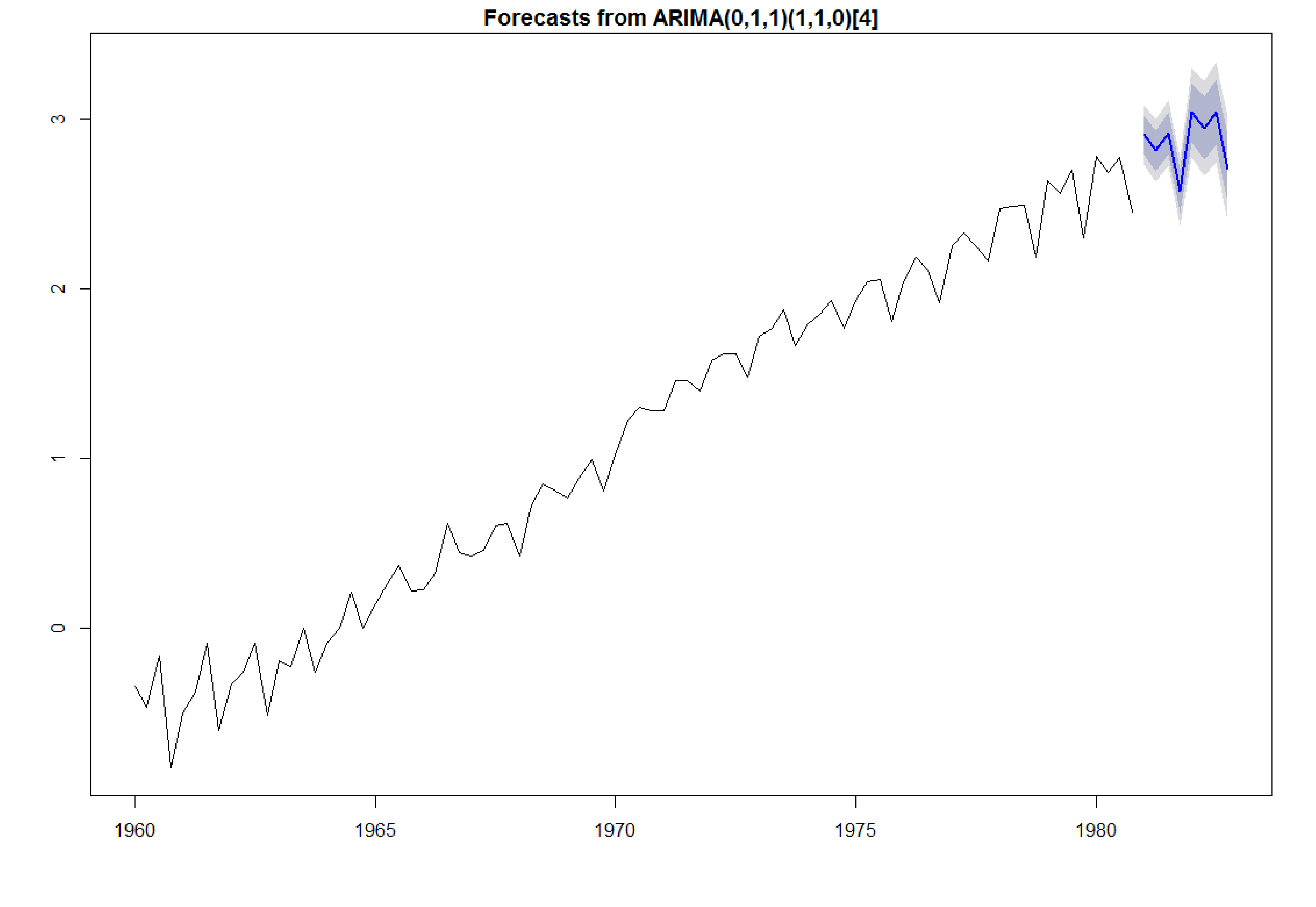
and

$$Z_t \sim Normal\ (0, 0.0079)$$

Forecast routines

- model<- arima(x=log(jj), order = c(0,1,1), seasonal = list(order=c(1,1,0), period=4))

- plot(forecast(model)) # ‘forecast’ package


Forecasts from ARIMA(0,1,1)(1,1,0)[4]
3
2
1
0
1960
1965
1970
1975
1980

forecast(model)

	Point for.	Lo 80	Hi 80	Lo 95	Hi 95
1981 Q1	2.910254	2.796250	3.024258	2.735900	3.084608
1981 Q2	2.817218	2.697507	2.936929	2.634135	3.000300
1981 Q3	2.920738	2.795580	3.045896	2.729325	3.112151
1981 Q4	2.574797	2.444419	2.705175	2.375401	2.774194
1982 Q1	3.041247	2.868176	3.214317	2.776559	3.305934
1982 Q2	2.946224	2.762623	3.129824	2.665431	3.227016
1982 Q3	3.044757	2.851198	3.238316	2.748735	3.340780
1982 Q4	2.706534	2.503505	2.909564	2.396028	3.017041

What We've Learned

- Fit SARIMA models to quarterly earnings of Johnson & Johnson share
- Forecast future values of examined time series