SARIMA processes

PRACTICAL TIME SERIES ANALYSIS

THISTLETON AND SADIGOV

Objectives

- Describe Seasonal ARIMA models
- Rewrite Seasonal ARIMA models using backshift and difference operators

ARIMA processes $\{X_t\}$

Let

$$Y_t = \nabla^d X_t$$

then

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots \theta_q Z_{t-q}$$

can be written as

$$\phi(B) Y_t = \theta(B) Z_t$$

where

$$\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$$

$$\phi\,(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

Box-Jenkins Seasonal ARIMA model

- Data might contain seasonal periodic component in addition to correlation with recent lags
- It repeats every s observations
- For a time series of monthly observations, X_t might depend on annual lags
- $X_{t-12}, X_{t-24}, \ldots$
- Quarterly data might have period of $s = 4$
- Seasonal ARIMA model

Pure Seasonal ARMA process

$ARMA(P,Q)_s$ has the form

$$\Phi_{\mathrm{P}}(B^s)X_t = \Theta_{\mathrm{Q}}(B^s)Z_t$$

where

$$\Phi_{\mathrm{P}}(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$$

and

$$\Theta_{\mathrm{Q}}(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs}$$

Stationarity and invertibility

Just like pure ARMA processes, for Seasonal ARMA process to be stationary and invertible, we need that the complex roots of the polynomials

$$\Phi_{\mathrm{P}}(z^{s})$$

and

$$\Theta_{\mathrm{Q}}(z^{s})$$

are outside of the unit circle.

Example 1

$Seasonal\ ARMA(1,0)_{12}$ has the form

$$(1 - \Phi_1 B^{12})X_t = Z_t$$

i.e.,

$$X_t = \Phi_1 \mathrm{X}_{\mathrm{t}-12} + \mathrm{Z}_{\mathrm{t}}$$

Example 2

$Seasonal\ ARMA(1,1)_{12}$ has the form

$$(1 - \Phi_1 B^{12})X_t = (1 + \Theta_1 B^{12})Z_t$$

i.e.,

$$X_t = \Phi_1 \mathrm{X}_{\mathrm{t}-12} + \mathrm{Z}_{\mathrm{t}} + \Theta_1 Z_{t-12}$$

Seasonal ARIMA process (SARIMA)

$SARIMA(p, d, q, P, D, Q)_s$ has the form

$$\Phi_P(B^s)\phi_p(B)(1-B^S)^D(1-B)^d X_t = \Theta_Q(B^s)\theta_q(B)Z_t$$

where

$$\theta_q(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$$

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs}$$

$$\phi_p\,(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$$

SARIMA models

$SARIMA(p, d, q, P, D, Q)_S$ has two parts:

Non-seasonal part (p, d, q) and seasonal parts $(P, D, Q)_S$.

1. p – order of non-seasonal AR terms
2. d – order of non-seasonal differencing
3. q – order of non-seasonal MA terms
4. P – order of seasonal AR (i.e., SAR) terms
5. D – order of seasonal differencing (i.e., power of $(1 - B^S)$)
6. Q – order of seasonal MA (i.e., SMA) terms

Seasonal Differencing

- $D = 1$

$$\nabla_S X_t = (1 - B^S)X_t = X_t - X_{t-S}$$

- $D = 2$

$$\nabla_S^2 X_t = (1 - B^S)^2 X_t = (1 - 2B^S + B^{2S})X_t = X_t - 2X_{t-S} + X_{t-2S}$$

Example 3- $SARIMA(1,0,0,1,0,1)_{12}$

$$(1-\phi_1 B)(1-\Phi_1 B^{12})X_t = (1+\Theta_1 B^{12})Z_t$$

$$(1-\phi_1 B-\Phi_1 B^{12}+\phi_1\Phi_1 B^{13})X_t = Z_t+\Phi_1 Z_{t-12}$$

Thus

$$X_t = \phi_1 X_{t-1}+\Phi_1 X_{t-12}-\phi_1\Phi_1 X_{t-13}+Z_t+\Phi_1 Z_{t-12}$$

Example 4 - $SARIMA(0,1,1,0,0,1)_4$

$$(1 - B)X_t = (1 + \Theta_1 B^4)(1 + \theta_1 B)Z_t$$

Then,

$$X_t - X_{t-1} = (1 + \theta_1 B + \Theta_1 B^4 + \theta_1 \Theta_1 B^5)Z_t$$

Thus

$$X_t = X_{t-1} + Z_t + \theta_1 Z_{t-1} + \Theta_1 Z_{t-4} + \theta_1 \Theta_1 Z_{t-5}$$

What We've Learned

- Describe seasonal, autoregressive, integrated, moving average models
- Rewrite seasonal, autoregressive, integrated, moving average models using backshift and difference operators

ACF of SARIMA processes

PRACTICAL TIME SERIES ANALYSIS

THISTLETON AND SADIGOV

Objectives

- Examine ACF of a SARIMA model in simulation
- Examine ACF of a SARIMA model in theory

Example - $SARIMA(0,0,1,0,0,1)_{12}$

$$X_t = (1 + \Theta_1 B^{12})(1 + \theta_1 B)Z_t$$

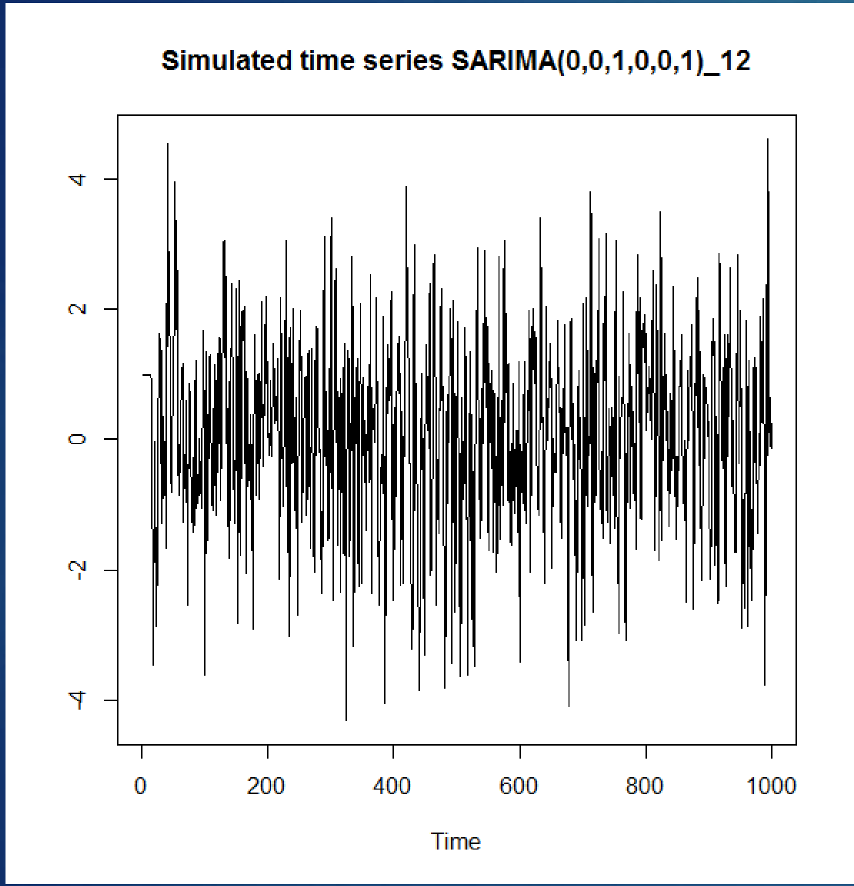
Thus

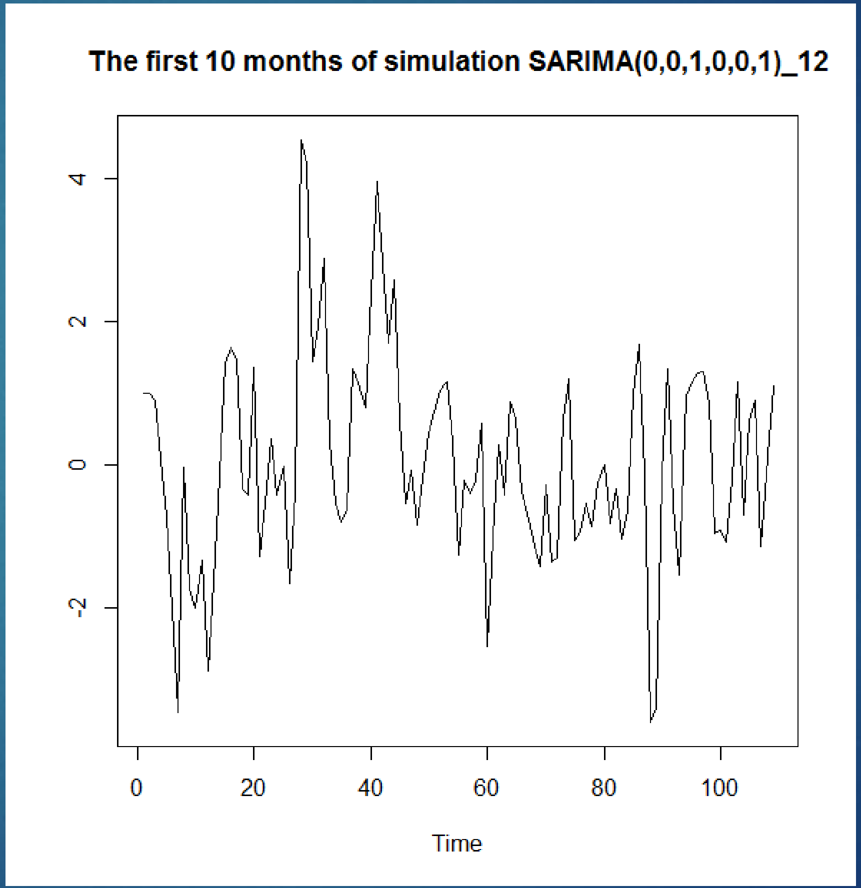
$$X_t = Z_t + \theta_1 Z_{t-1} + \Theta_1 Z_{t-12} + \theta_1 \Theta_1 Z_{t-13}$$

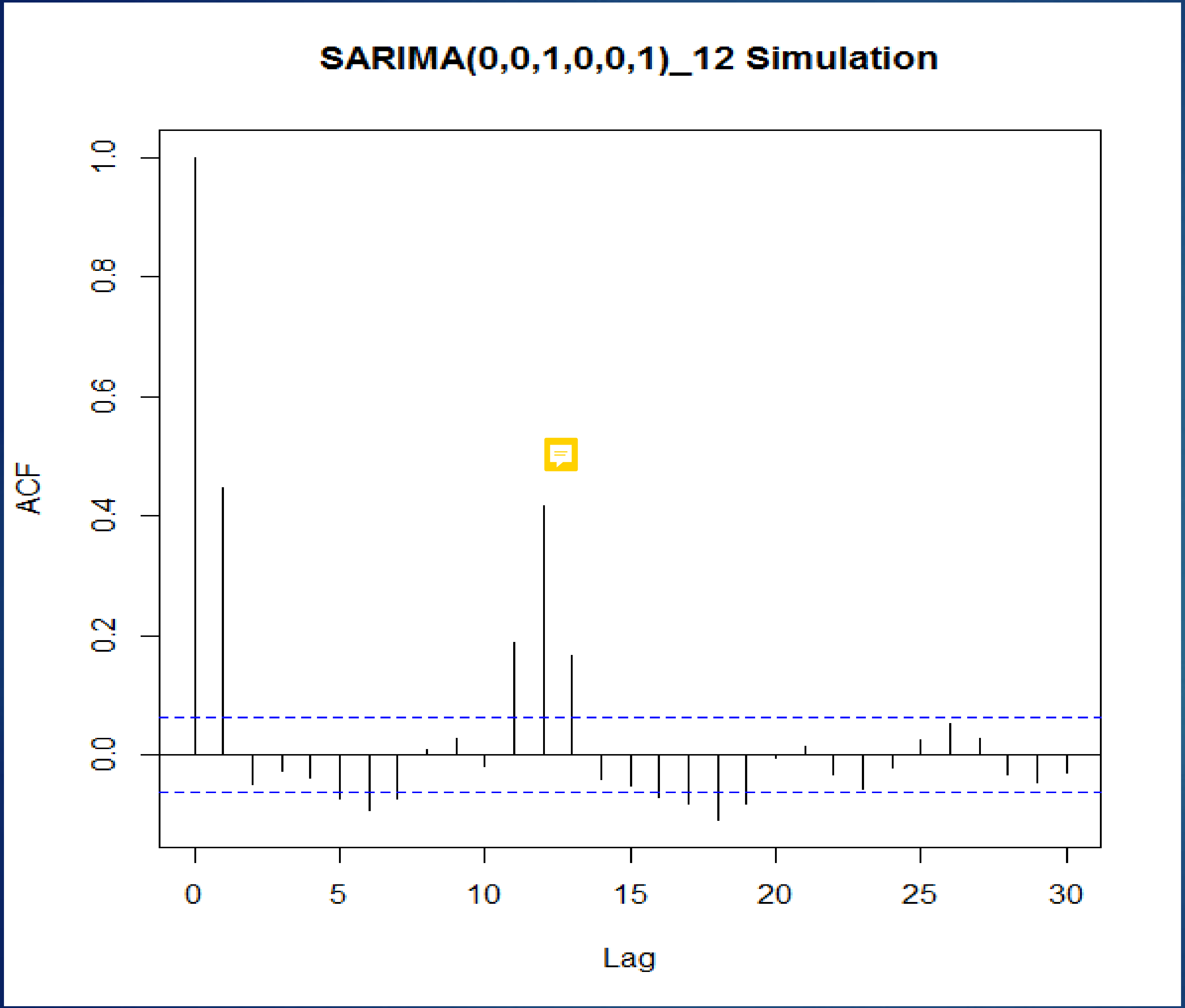
Choose $\theta_1 = 0.7, \Theta_1 = 0.6$, then

$$X_t = Z_t + 0.7\, Z_{t-1} + 0.6\, Z_{t-12} + 0.42\, Z_{t-13}$$

Simulation


Simulated time series SARIMA(0,0,1,0,0,1)_12
4
2
0
-2
-4
0
200
400
600
800
1000
Time


The first 10 months of simulation SARIMA(0,0,1,0,0,1)_12
4
2
0
-2
0
20
40
60
80
100
Time


SARIMA(0,0,1,0,0,1)_12 Simulation
ACF
Lag
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30

Example - $SARIMA(0,0,1,0,0,1)_{12}$

$$X_t = (1 + \Theta_1 B^{12})(1 + \theta_1 B)Z_t$$

Thus

$$X_t = Z_t + \theta_1 Z_{t-1} + \Theta_1 Z_{t-12} + \theta_1 \Theta_1 Z_{t-13}$$

Autocovariance function: $\gamma(k)$

$$\gamma(0) = Cov(X_t, X_t) = Var(X_t)$$

$$X_t = Z_t + \theta_1 Z_{t-1} + \Theta_1 Z_{t-12} + \theta_1 \Theta_1 Z_{t-13}$$

$$Var(X_t) = \sigma_Z^2 + \theta_1^2 \sigma_Z^2 + \Theta_1^2 \sigma_Z^2 + \theta_1^2 \Theta_1^2 \sigma_Z^2$$

$$\gamma(0) = (1 + \theta_1^2)(1 + \Theta_1^2)\sigma_Z^2$$

$\gamma(1)$

$$\gamma(1) = Cov(X_t, X_{t-1})$$

$$X_t = Z_t + \theta_1 Z_{t-1} + \Theta_1 Z_{t-12} + \theta_1 \Theta_1 Z_{t-13}$$

$$X_{t-1} = Z_{t-1} + \theta_1 Z_{t-2} + \Theta_1 Z_{t-13} + \theta_1 \Theta_1 Z_{t-14}$$

$$\gamma(1) = \theta_1 \sigma_Z^2 + \theta_1 \Theta_1^2 \sigma_Z^2$$

$$\gamma(1) = \theta_1 (1 + \Theta_1^2) \sigma_Z^2$$

ACF: $\rho(1)$

$$\gamma(1) = \theta_1(1 + \Theta_1^2)\sigma_Z^2$$

$$\gamma(0) = (1 + \theta_1^2)(1 + \Theta_1^2)\sigma_Z^2$$

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{\theta_1}{1 + \theta_1^2} \leq \frac{1}{2}$$

Since $(\theta_1 - 1)^2 \geq 0$

$\gamma(2)$

$$\gamma(2) = Cov(X_t, X_{t-2})$$

$$X_t = Z_t + \theta_1 Z_{t-1} + \Theta_1 Z_{t-12} + \theta_1 \Theta_1 Z_{t-13}$$

$$X_{t-2} = Z_{t-2} + \theta_1 Z_{t-3} + \Theta_1 Z_{t-14} + \theta_1 \Theta_1 Z_{t-15}$$

$$\gamma(2) = 0$$

since $Z_t's$ are independent.

Thus

$$\rho(2) = 0$$

ACF

$$\rho(i) = 0$$

when $i = 2,3,\ldots,10.$

$\gamma(11)$, $\rho(11)$

$$\gamma(11) = Cov(X_t, X_{t-11})$$

$$X_t = Z_t + \theta_1 Z_{t-1} + \Theta_1 Z_{t-12} + \theta_1 \Theta_1 Z_{t-13}$$

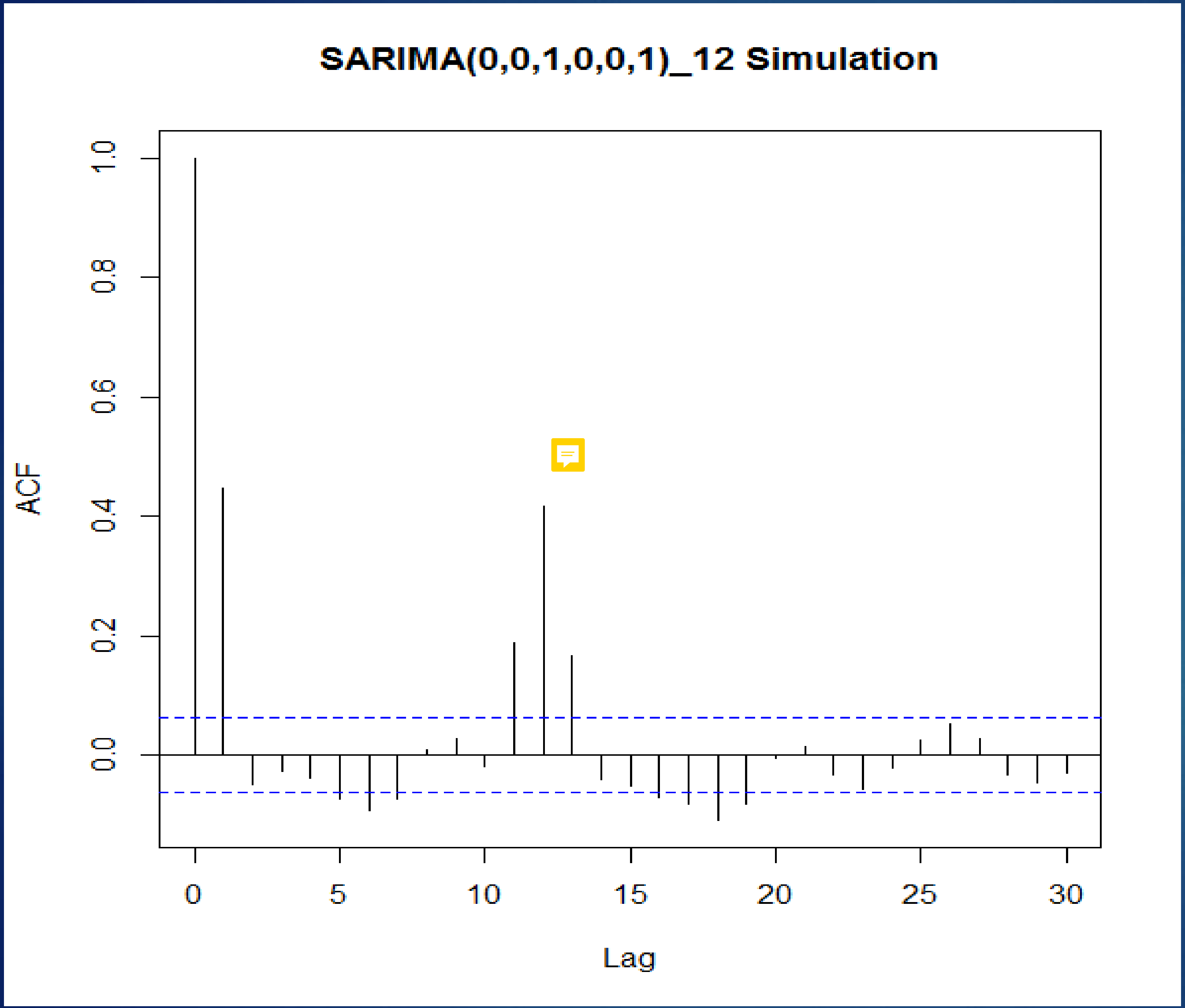
$$X_{t-11} = Z_{t-11} + \theta_1 Z_{t-12} + \Theta_1 Z_{t-23} + \theta_1 \Theta_1 Z_{t-24}$$

$$\gamma(11) = \theta_1 \Theta_1 \sigma_Z^2$$

$$\rho(11) = \frac{\gamma(11)}{\gamma(0)} = \frac{\theta_1 \Theta_1}{(1+\theta_1^2)(1+\Theta_1^2)} \neq 0$$

But

$$0 < \rho(11) \leq \frac{1}{4}$$



SARIMA(0,0,1,0,0,1)_12 Simulation
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

What We've Learned

- ACF of a SARIMA model in simulation
- ACF of a SARIMA model in theory

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

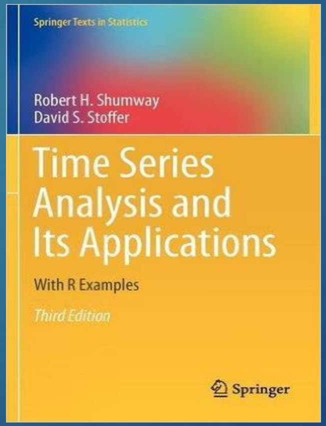
Shumway, R.H. and Stoffer, D.S. (2000)

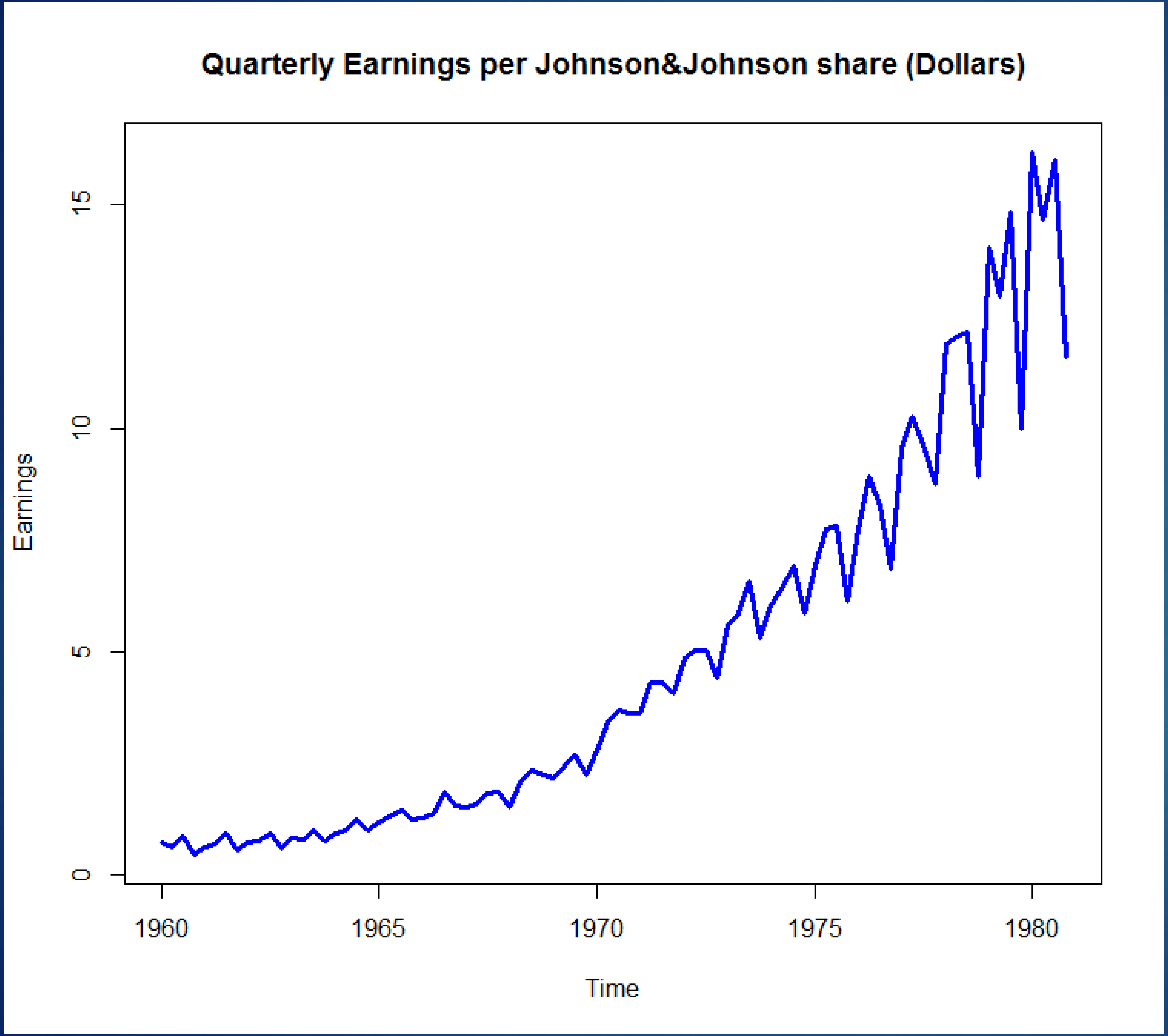
Time Series Analysis and its Applications

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

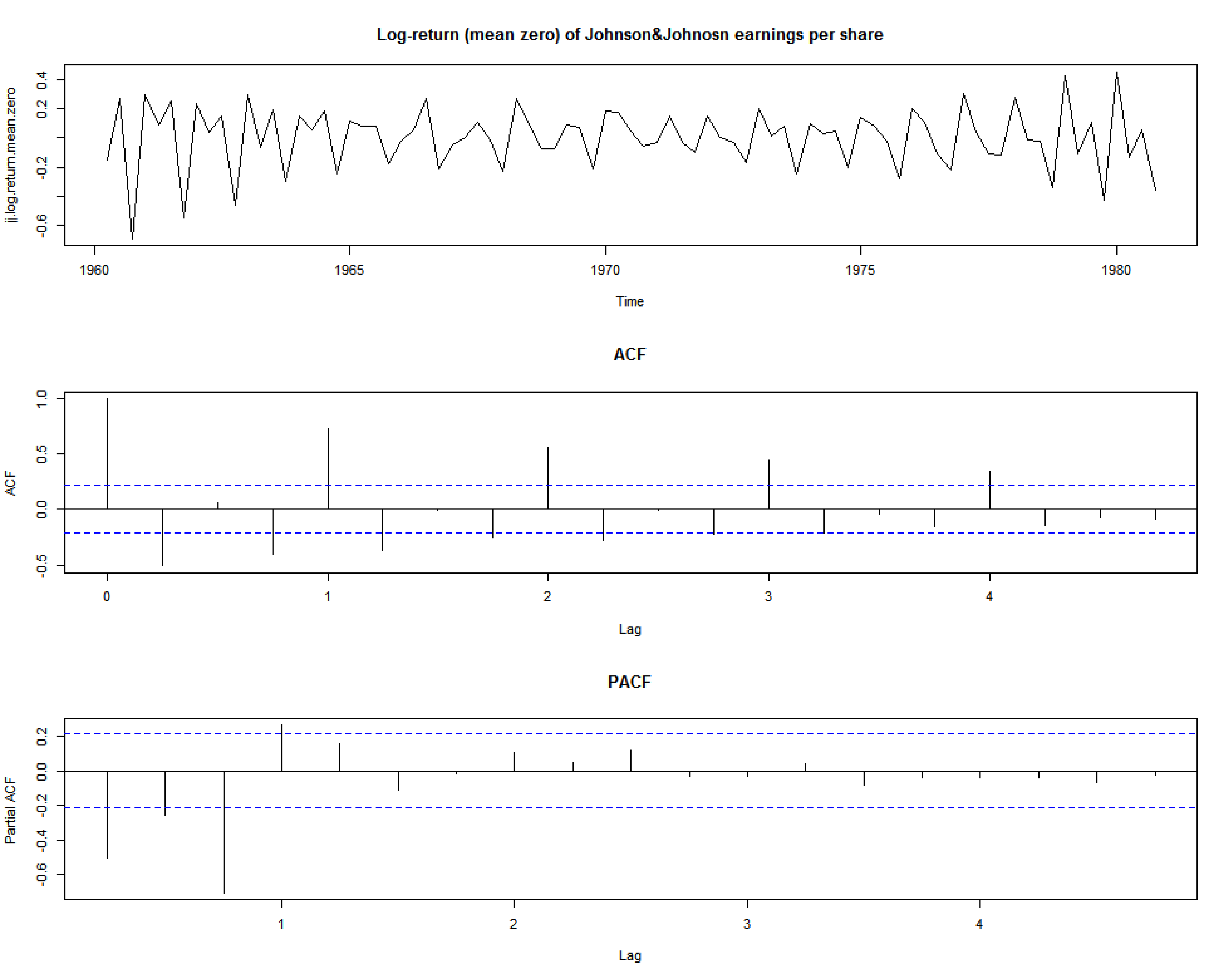
Log-return a time series $\{X_t\}$

is defined as

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

In R,

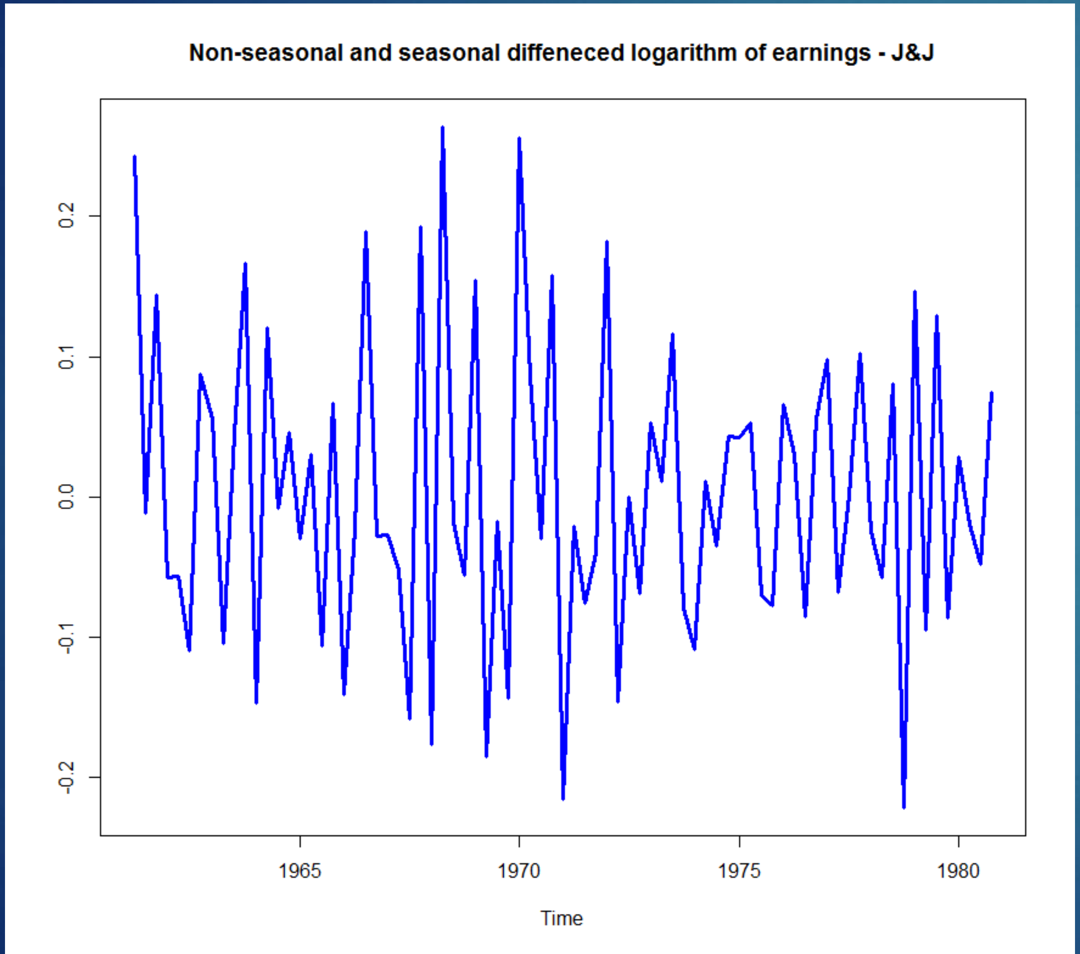
$$diff(\log(\quad))$$



Log-return (mean zero) of Johnson&Johnosn earnings per share
jj.log.return.mean.zero
Time
ACF
Lag
PACF
Partial ACF
Lag

Seasonal differencing D=1

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



Ljung-Box test

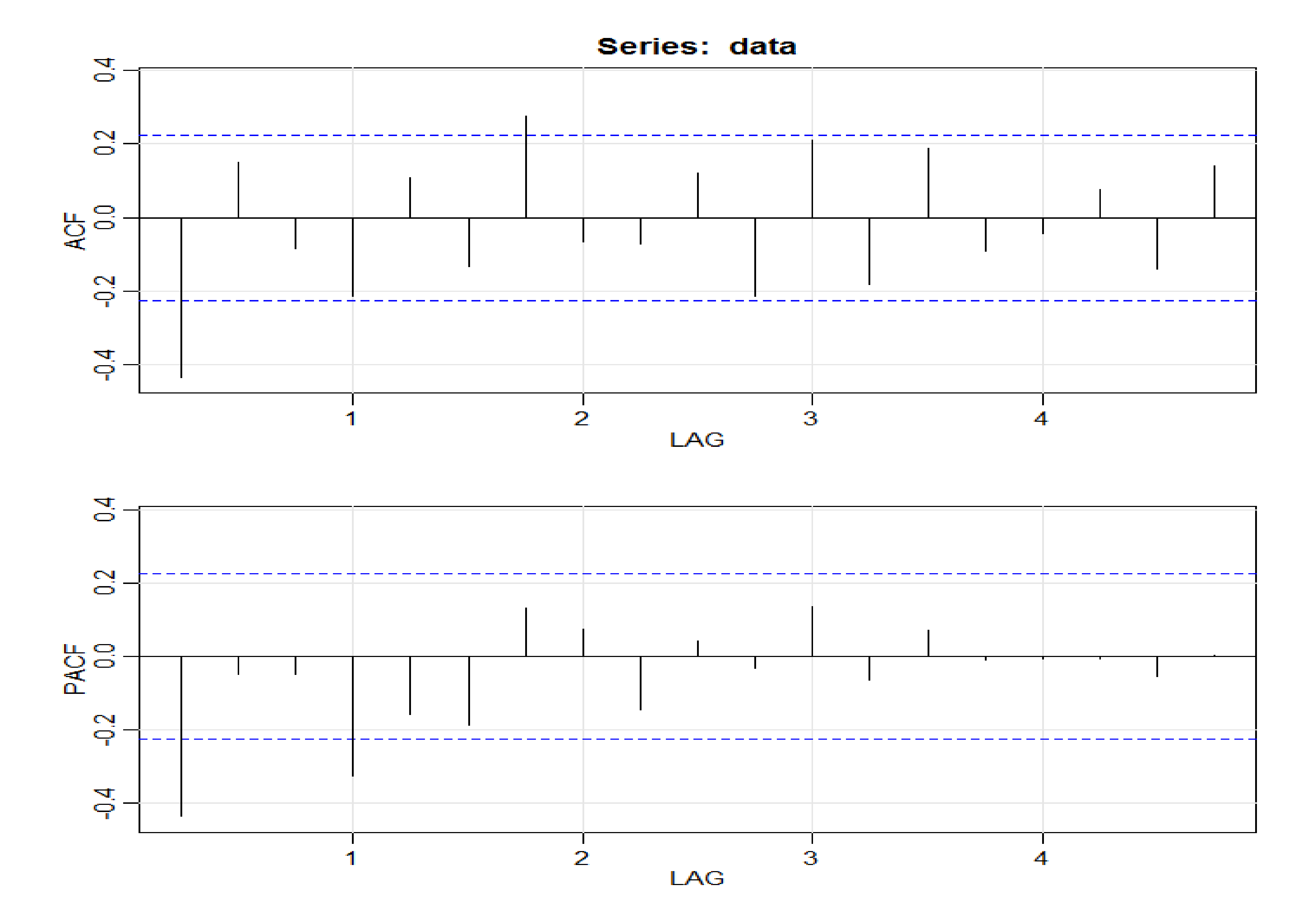
- `Box.test(data, lag=log(length(data)))`
- p-value:

$$0.0004658$$

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

$$X_t = Earnings$$

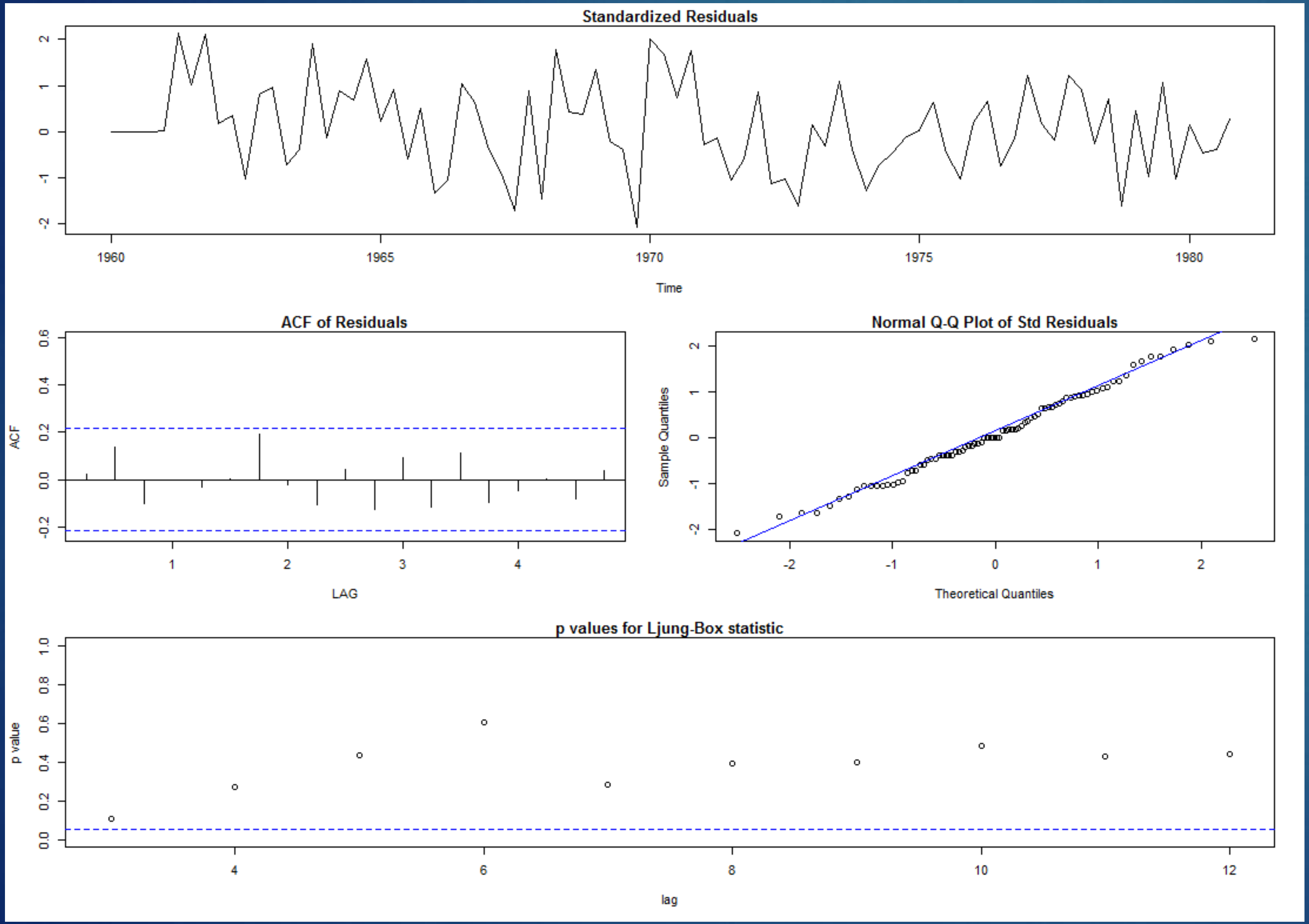
$$Y_t = \log(X_t)$$

	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

$$Y_t = \log(X_t)$$

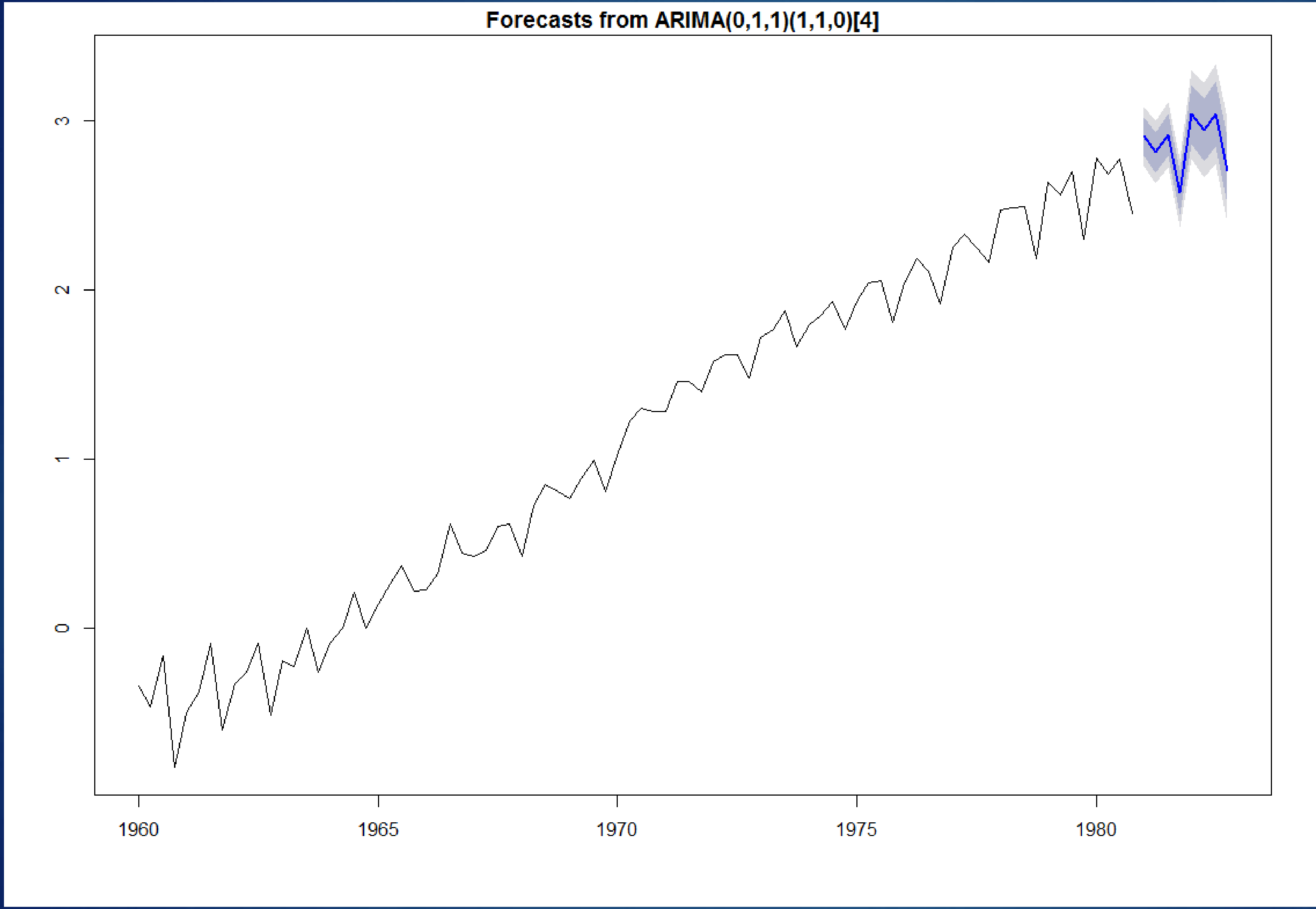
and

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

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]
0
1
2
3
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

SARIMA fitting: Milk production

PRACTICAL TIME SERIES ANALYSIS

THISTLETON AND SADIGOV

Objectives

- Fit SARIMA models to Milk production data from TSDL
- Forecast future values of examined time series

Modeling

- Time plot
- Transformation
- Differencing (seasonal or non-seasonal)
- 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$(p, d, q, P, D, Q)_S$

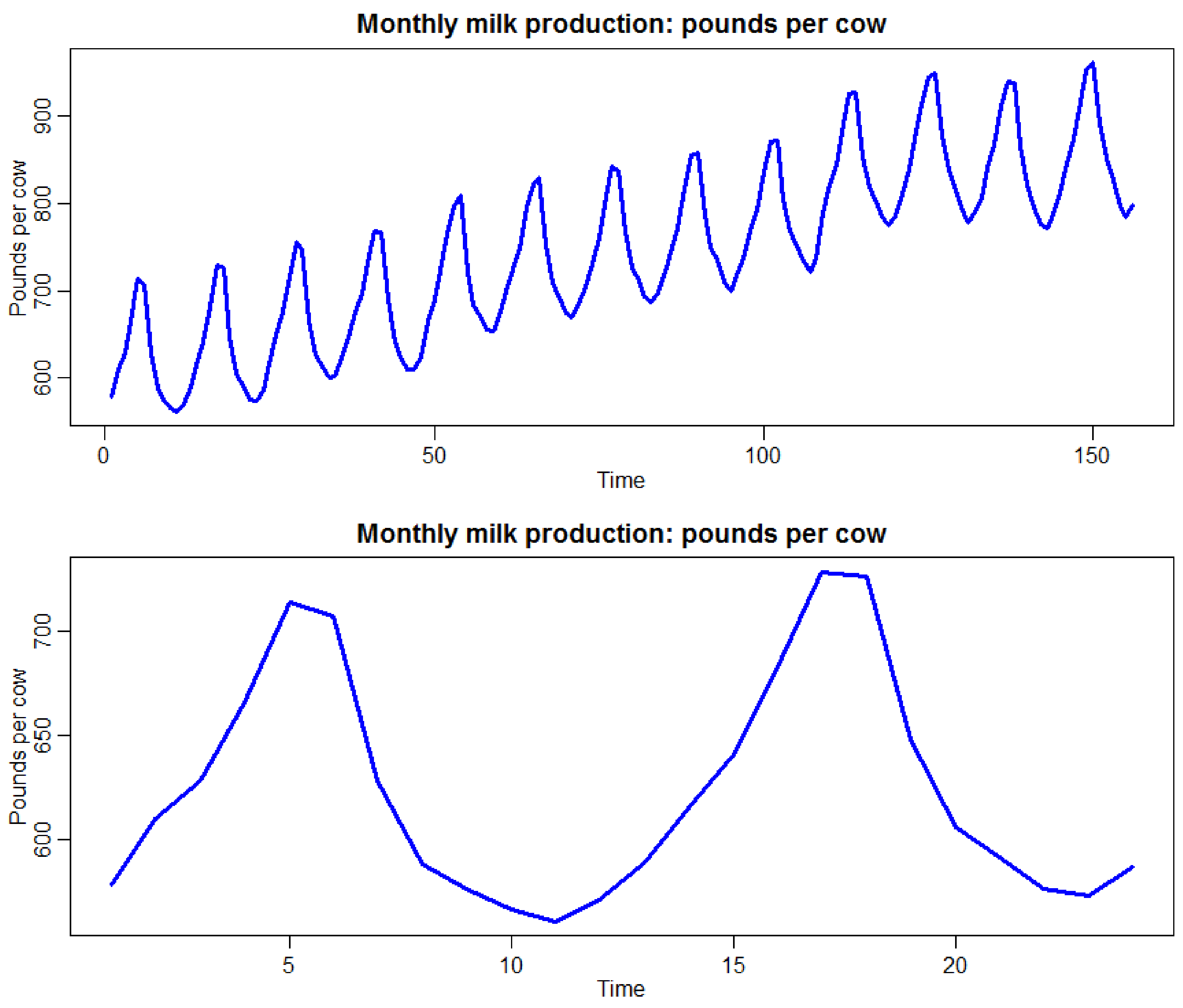
$$p + d + q + P + D + Q \leq 6$$

Time Series Data Library

- TSDL
- Created by Rob Hyndman
- Professor of Statistics
- Monash University, Australia
- https://datamarket.com/data/list/?q=provider%3Atsdl

Monthly milk production : Agriculture

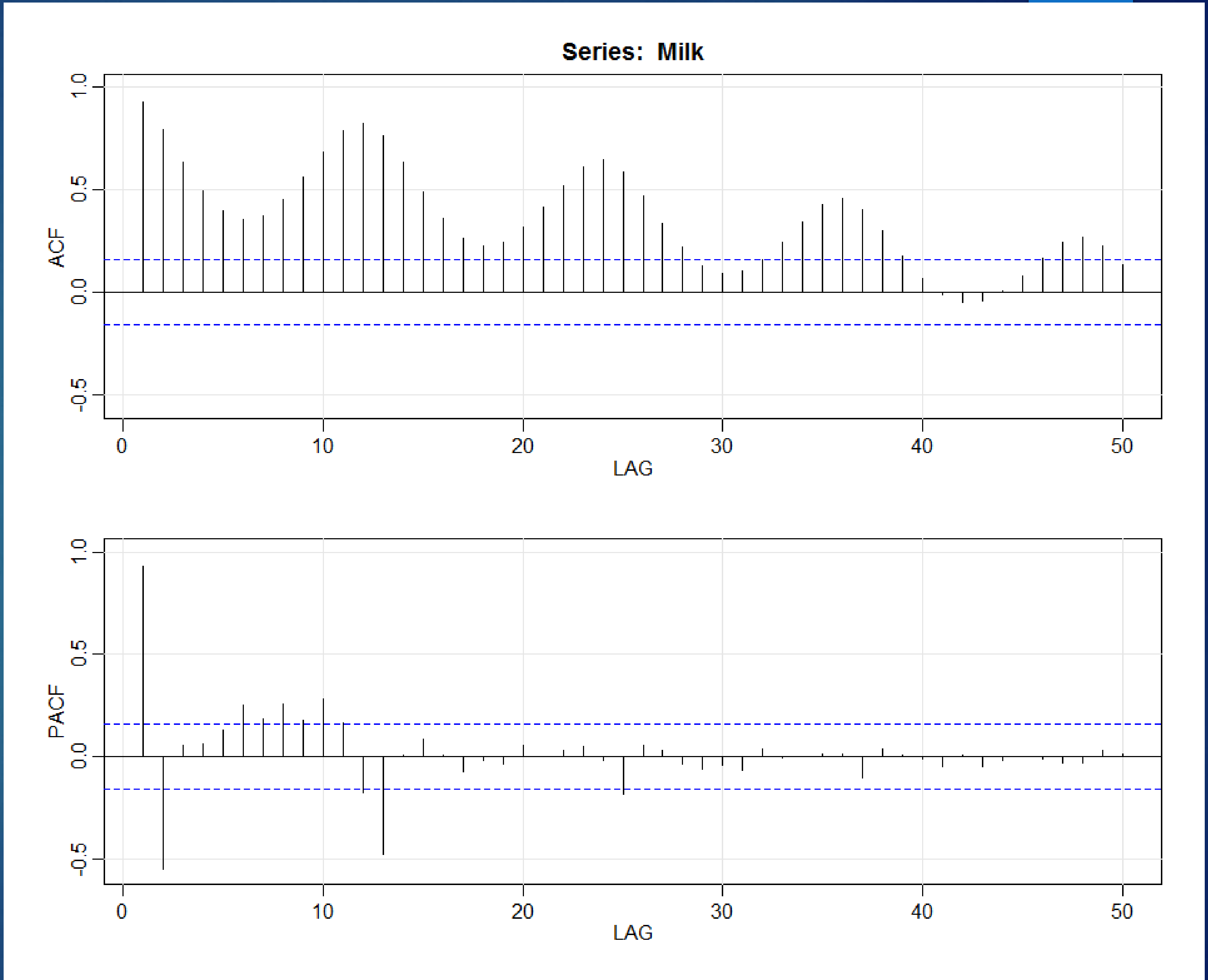
- https://datamarket.com/data/set/22sn/monthly-milk-production-pounds-per-cow-jan-62-dec-75-adjusted-for-month-length#!ds=22sn&display=line
- Monthly milk production
- Pounds per cow
- January 1962 – December 1975
- Agriculture, Source: Cryer (1986)



Monthly milk production: pounds per cow
Pounds per cow
600
700
800
900
0
50
100
150
Time
Monthly milk production: pounds per cow
Pounds per cow
600
650
700
5
10
15
20
Time

ACF

PACF



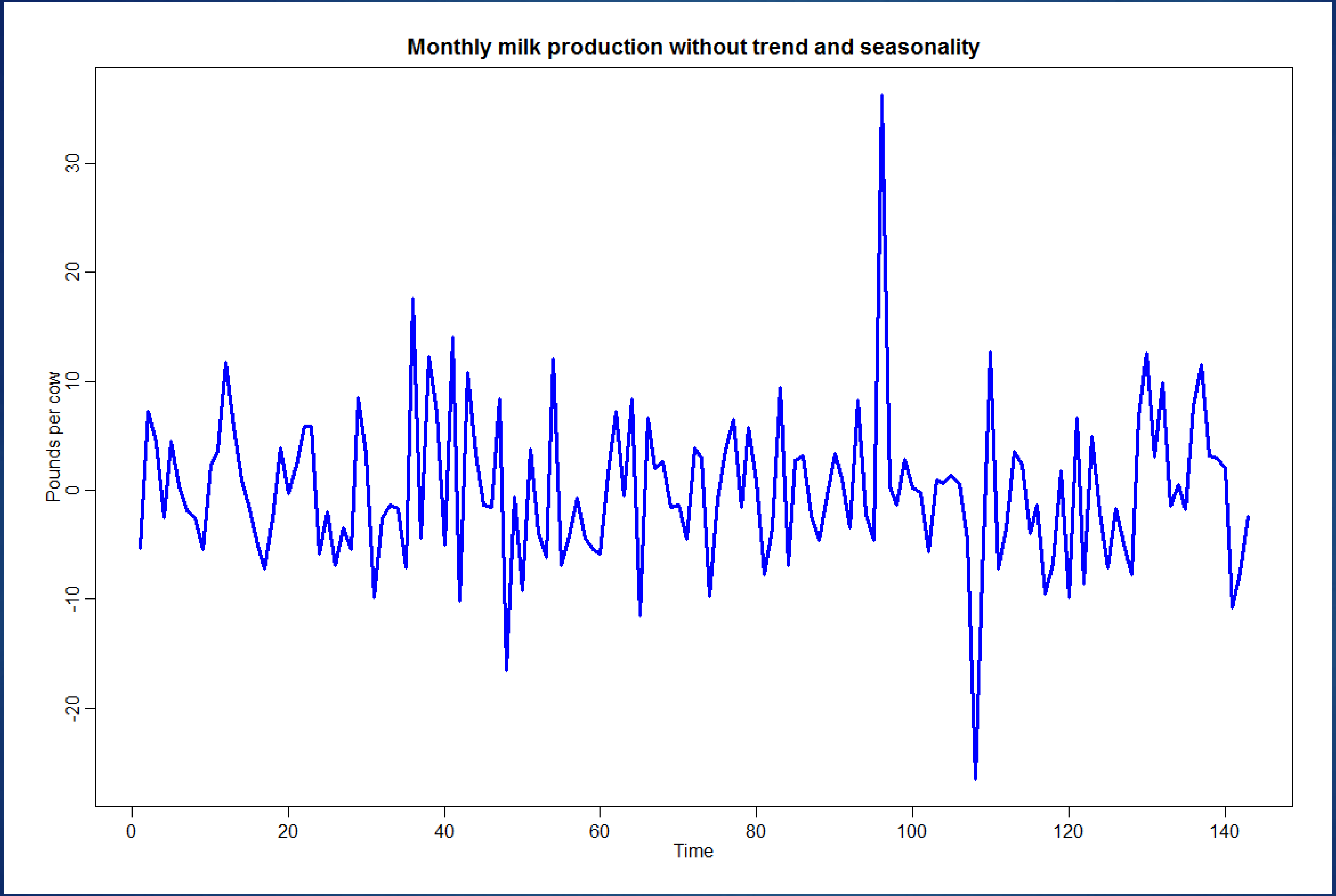
Series: Milk
ACF
1.0
0.5
0.0
-0.5
0
10
20
30
40
50
LAG
PACF
1.0
0.5
0.0
-0.5
0
10
20
30
40
50
LAG

Non-seasonal and seasonal differencing

$$d = 1$$

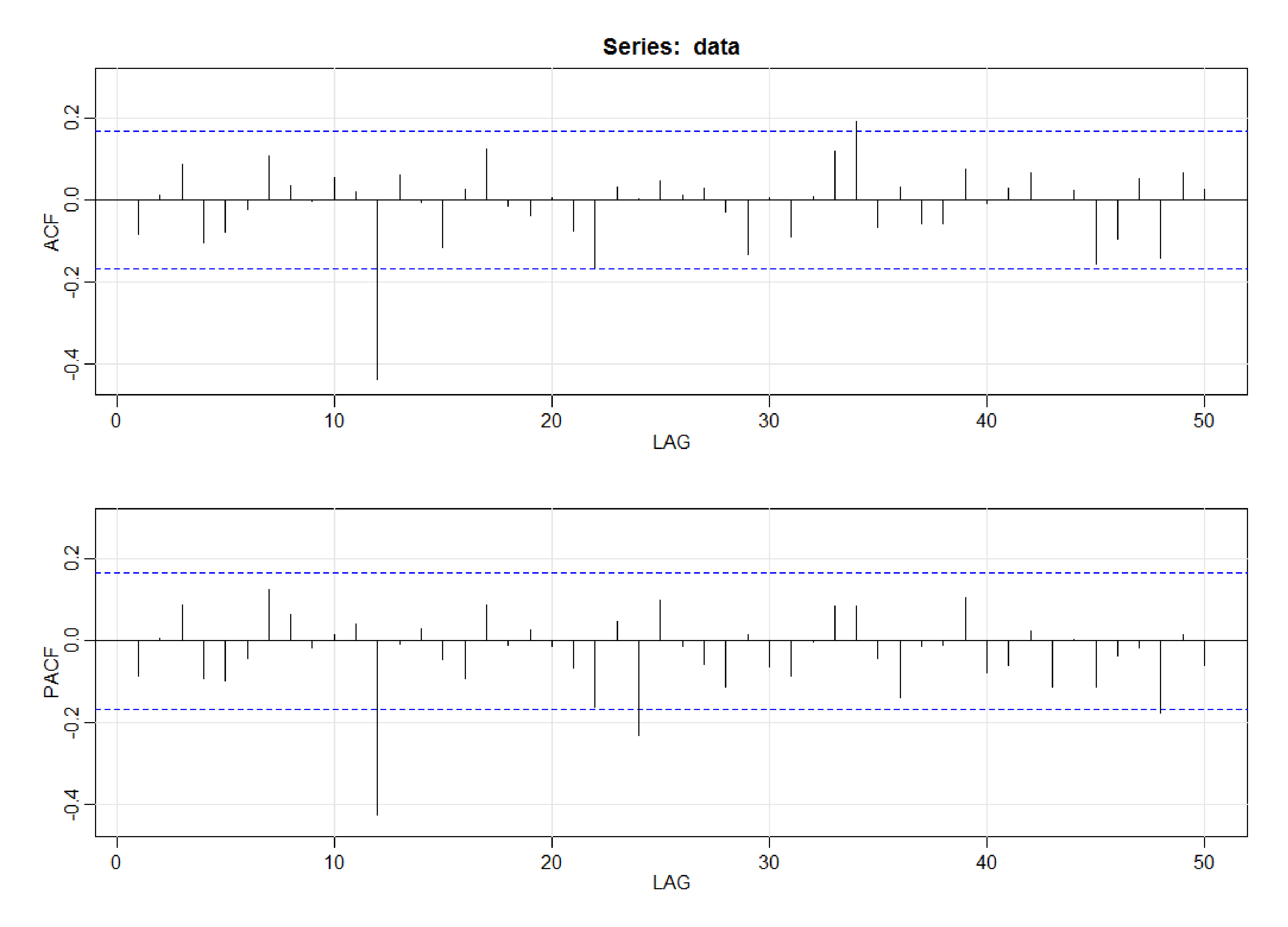
$$D = 1$$

$$diff(diff(milk), 12)$$


Monthly milk production without trend and seasonality
Pounds per cow
30
20
10
0
-10
-20
0
20
40
60
80
100
120
140
Time

ACF

PACF



Series: data
ACF
0.2
0.0
-0.2
-0.4
0
10
20
30
40
50
LAG
PACF
0.2
0.0
-0.2
-0.4
0
10
20
30
40
50
LAG

Order specification

- ACF → $q = 0\ ; Q = 0, 1, 2, 3$

- PACF → $p = 0; \mathrm{P} = 0\ , 1, 2$

0 1 0 0 1 0 12 AIC= 968.3966 SSE= 7213.013 p-VALUE= 0.4393367

0 1 0 0 1 1 12 AIC= 923.3288 SSE= 4933.349 p-VALUE= 0.6493728

0 1 0 0 1 2 12 AIC= 925.3072 SSE= 4931.398 p-VALUE= 0.6529998

0 1 0 0 1 3 12 AIC= 927.2329 SSE= 4925.911 p-VALUE= 0.6640233

0 1 0 1 1 0 12 AIC= 938.6402 SSE= 5668.197 p-VALUE= 0.493531

0 1 0 1 1 1 12 AIC= 925.3063 SSE= 4931.428 p-VALUE= 0.6531856

0 1 0 1 1 2 12 AIC= 927.3036 SSE= 4931.135 p-VALUE= 0.6537708

0 1 0 1 1 3 12 AIC= 929.2146 SSE= 4924.747 p-VALUE= 0.6627108

0 1 0 2 1 0 12 AIC= 932.6438 SSE= 5308.012 p-VALUE= 0.6004804

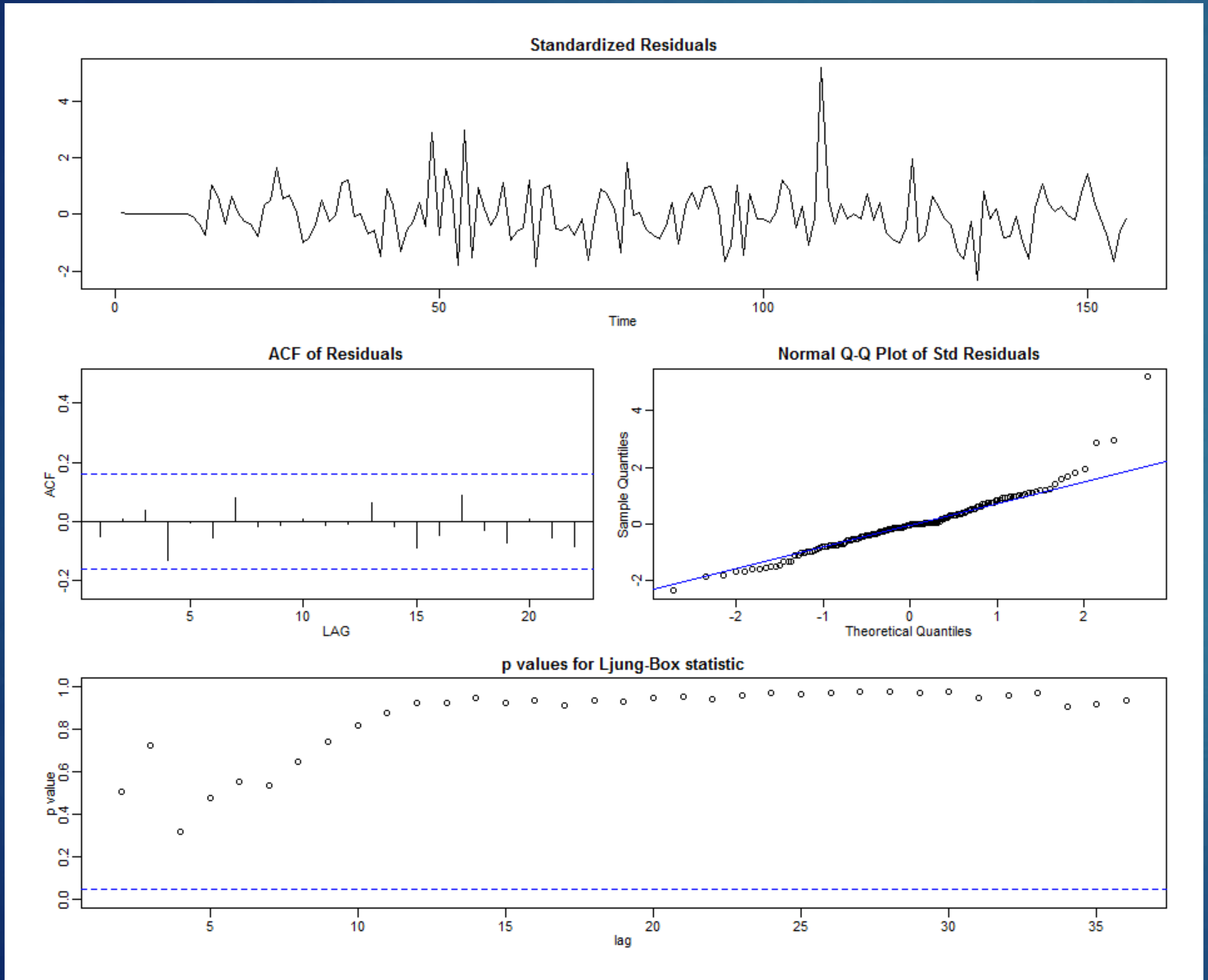
0 1 0 2 1 1 12 AIC= 927.2797 SSE= 4929.733 p-VALUE= 0.657349

0 1 0 2 1 2 12 AIC= 926.8053 **SSE= 4618.498** p-VALUE= 0.6826743

$SARIMA(0,1,0,0,1,1)_{12}$

	Estimate	SE	t.value	p.value
sma1	-0.6750	0.0752	-8.9785	0.0000

Residual analysis



Model – SARIMA$(0,1,0,0,1,1)_{12}$

$$X_t = Milk\ production\ pounds\ per\ cow$$

$$(1 - B)(1 - B^{12})X_t = (1 + \Theta B^{12})Z_t$$

$$X_t = X_{t-1} + X_{t-12} - X_{t-13} + Z_t + \Theta\, Z_{t-12}$$

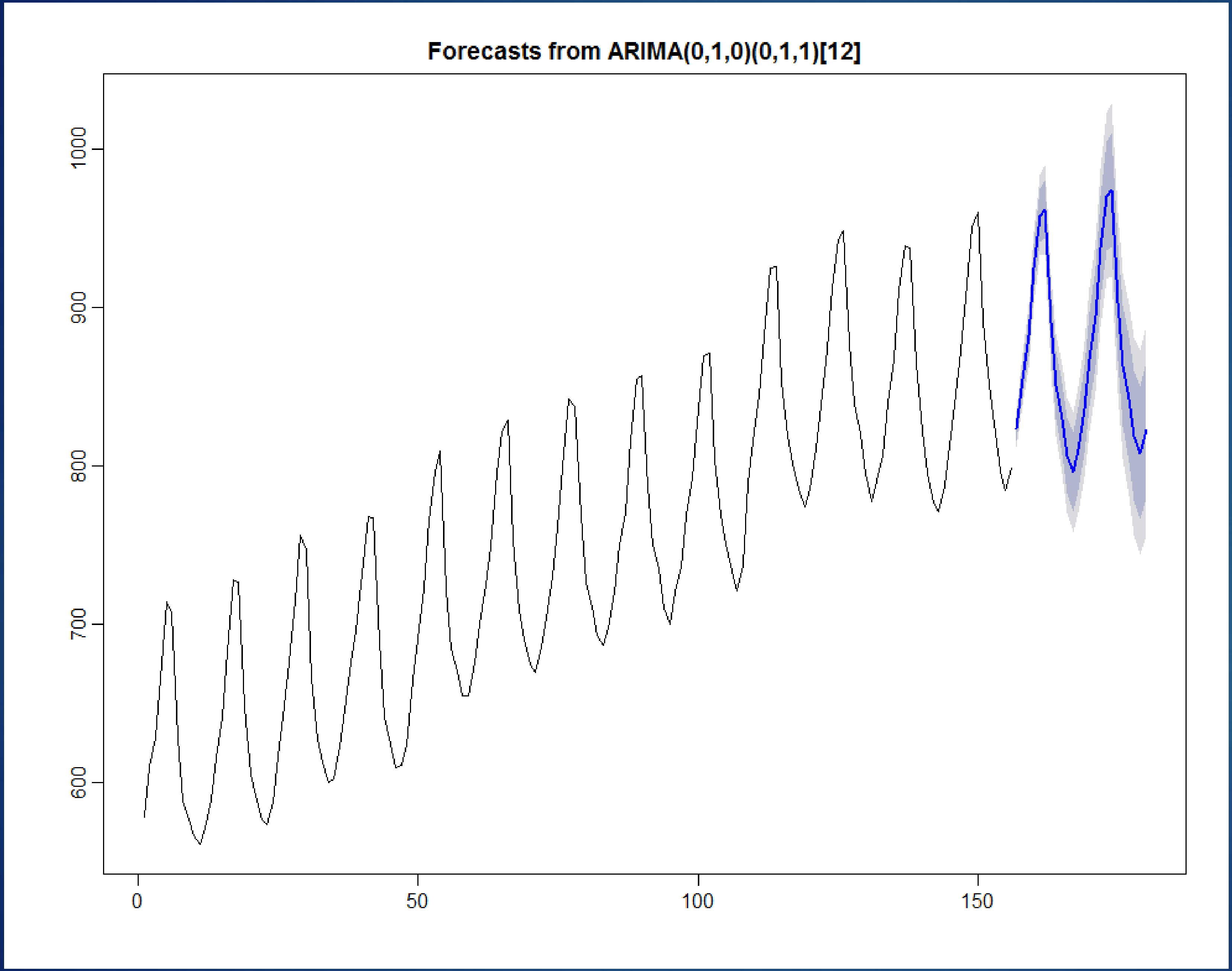
$$\widehat{\Theta} = -0.6750$$

Model – cont.

$$X_t = X_{t-1} + X_{t-12} - X_{t-13} + Z_t - 0.6750\, Z_{t-12}$$

where

$$Z_t \sim Normal\,(0, 34.47)$$


Forecasts from ARIMA(0,1,0)(0,1,1)[12]
1000
900
800
700
600
0
50
100
150

forecast(model)

	Pt. for.	Lo 80	Hi 80	Lo 95	Hi 95
157	823.3978	815.8740	830.9216	811.8911	834.9045
158	854.9196	844.2793	865.5598	838.6467	871.1925
159	882.1923	869.1607	895.2239	862.2622	902.1224
160	925.2390	910.1914	940.2866	902.2257	948.2523
161	958.4461	941.6225	975.2698	932.7165	984.1757
162	962.2105	943.7811	980.6399	934.0252	990.3959
163	890.9973	871.0912	910.9033	860.5536	921.4409
164	851.3336	830.0531	872.6140	818.7879	883.8792
165	829.7513	807.1800	852.3226	795.2314	864.2711
166	806.7802	782.9880	830.5725	770.3931	843.1673
167	795.9513	770.9978	820.9048	757.7882	834.1144
168	810.5435	784.4804	836.6066	770.6834	850.4036

What We've Learned

- Fit SARIMA models to Milk production data from TSDL
- Forecast future values of examined time series

SARIMA fitting: Sales at a souvenir shop

PRACTICAL TIME SERIES ANALYSIS

THISTLETON AND SADIGOV

Objectives

- Fit SARIMA models to dataset about sales at a souvenir shop from TSDL
- Forecast future values of examined time series

Modeling

- Time plot
- Transformation
- Differencing (seasonal or non-seasonal)
- 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$(p, d, q, P, D, Q)_S$

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

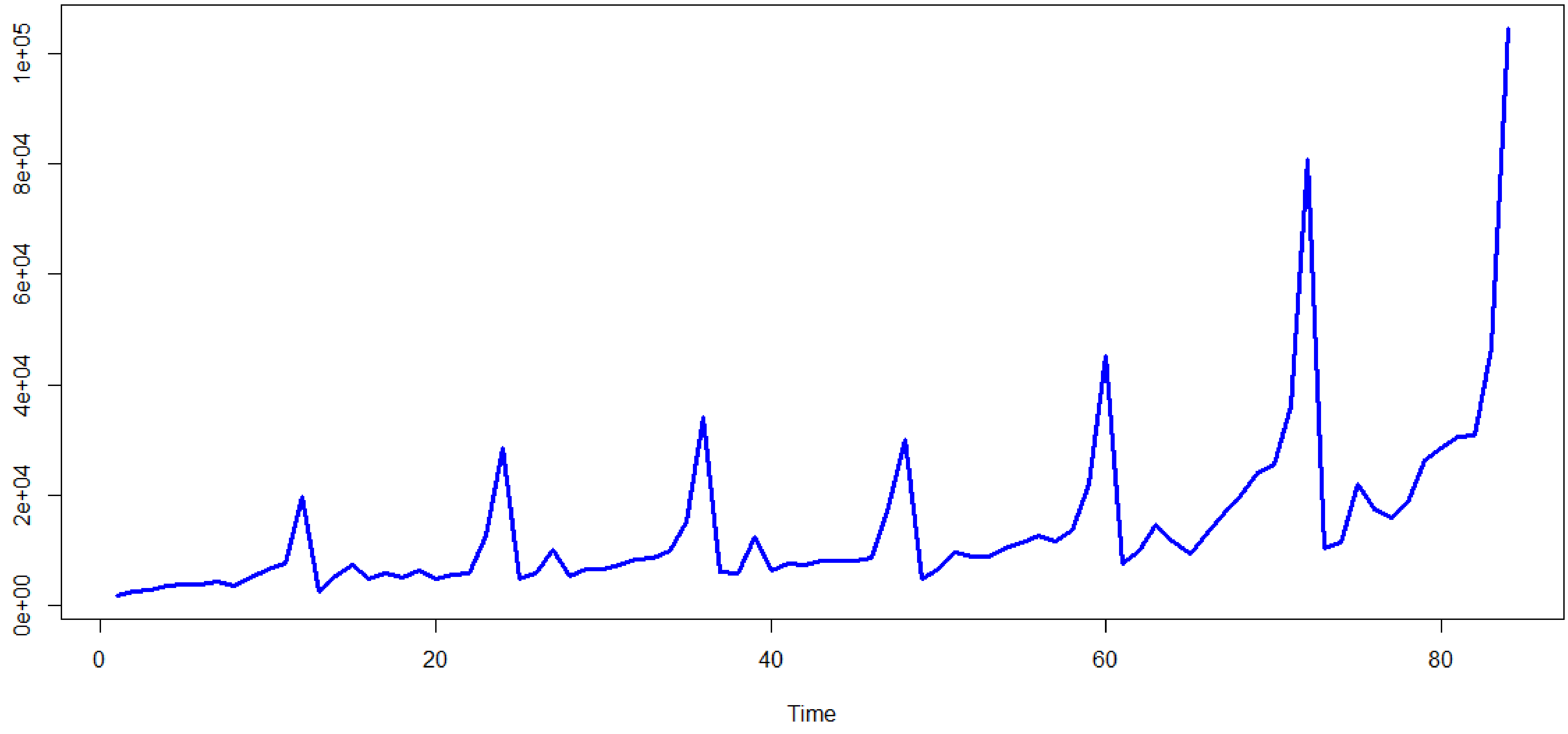
Time Series Data Library

- TSDL
- Created by Rob Hyndman
- Professor of Statistics
- Monash University, Australia
- https://datamarket.com/data/list/?q=provider%3Atsdl

Monthly sales for a souvenir shop: Sales

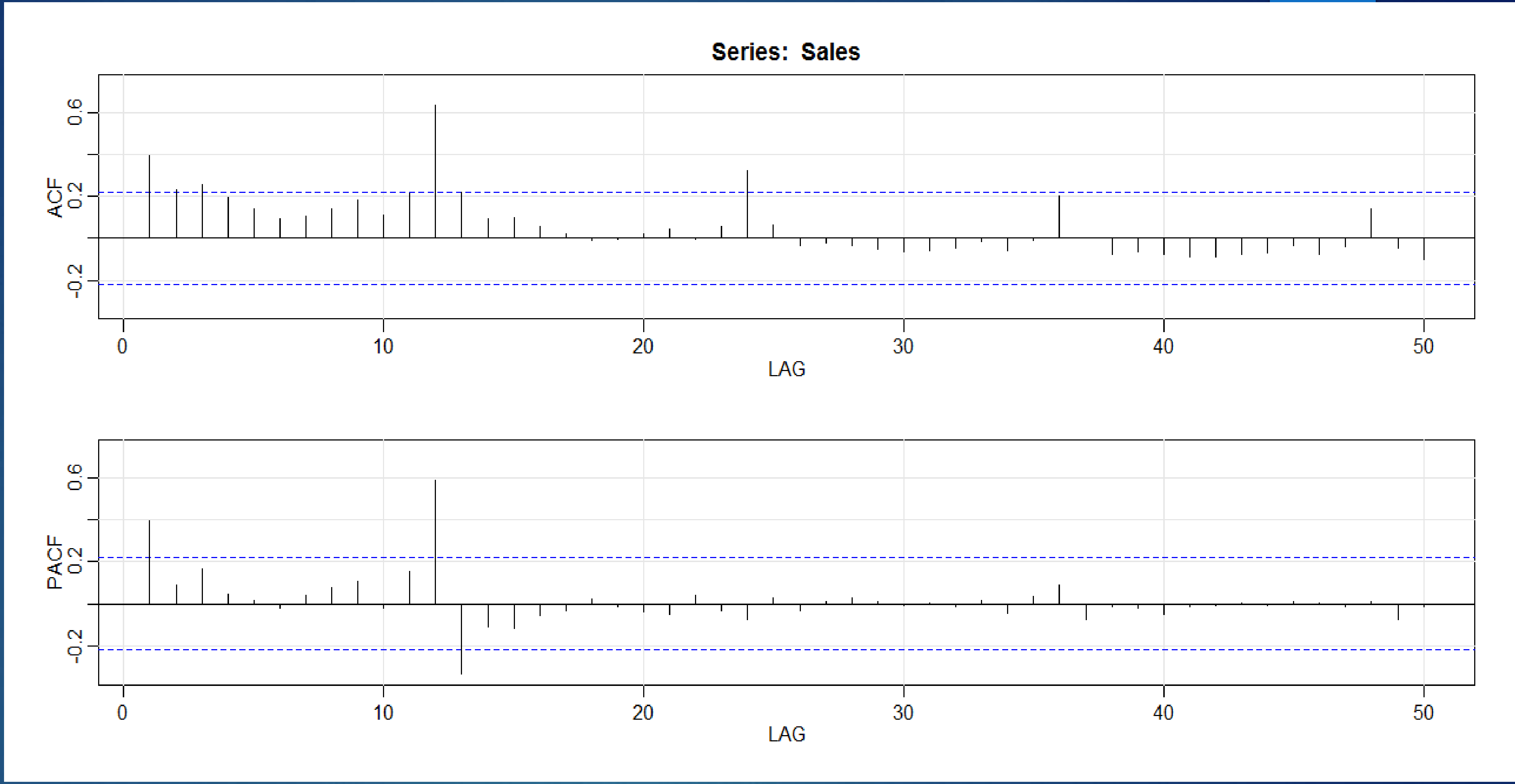
- https://datamarket.com/data/set/22mh/monthly-sales-for-a-souvenir-shop-on-the-wharf-at-a-beach-resort-town-in-queensland-australia-jan-1987-dec-1993#!ds=22mh&display=line
- Sales for a souvenir shop in Queensland, Australia
- January 1987 – December 1993
- Sales, Source: Makridakis, Wheelwright and Hyndman (1998)

Monthly sales for a souvenir shop in Queensland, Australia. Jan 1987-Dec 1993

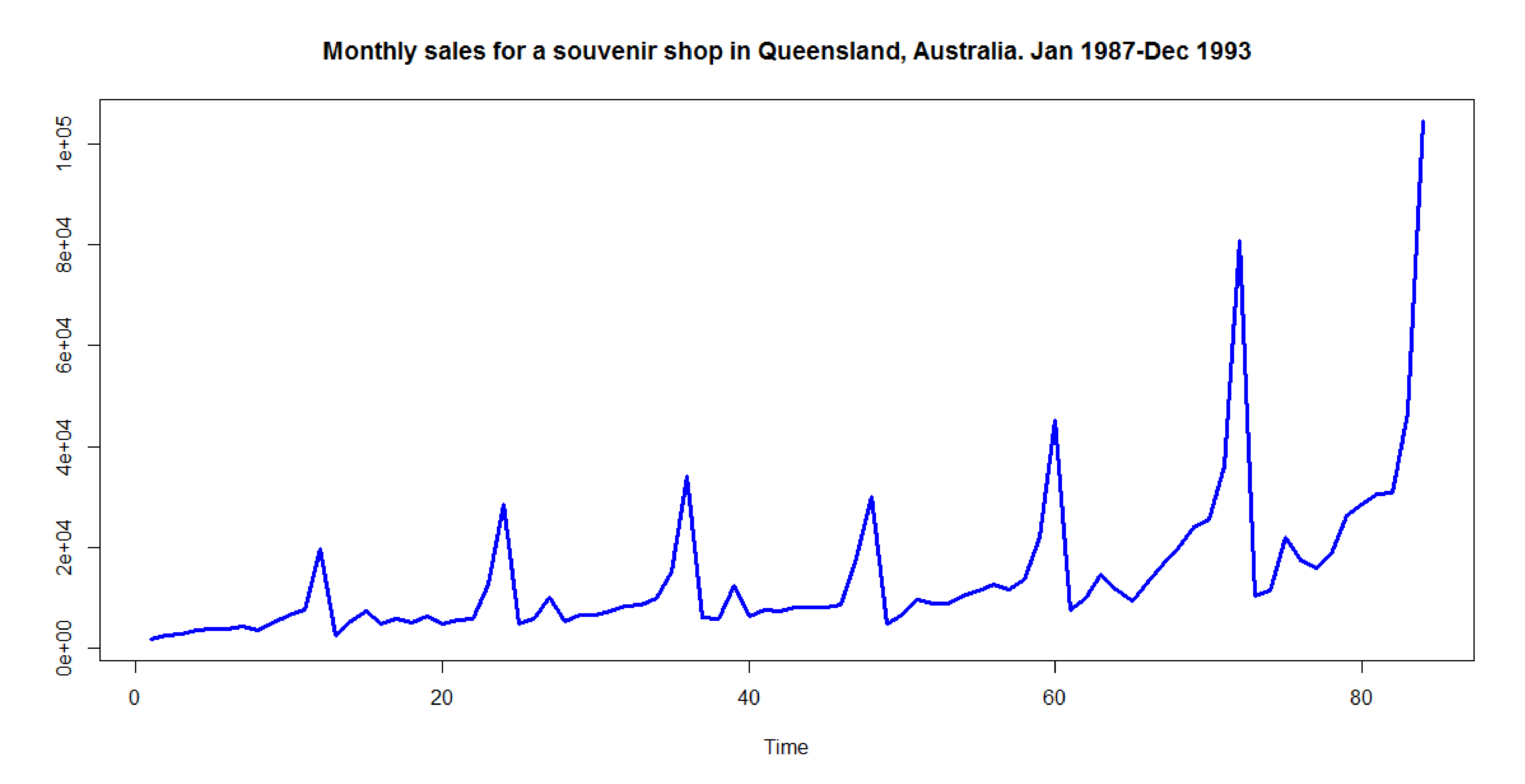


ACF

PACF



Series: Sales
ACF
0.6
0.2
-0.2
0
10
20
30
40
50
LAG
PACF
0.6
0.2
-0.2
0
10
20
30
40
50
LAG

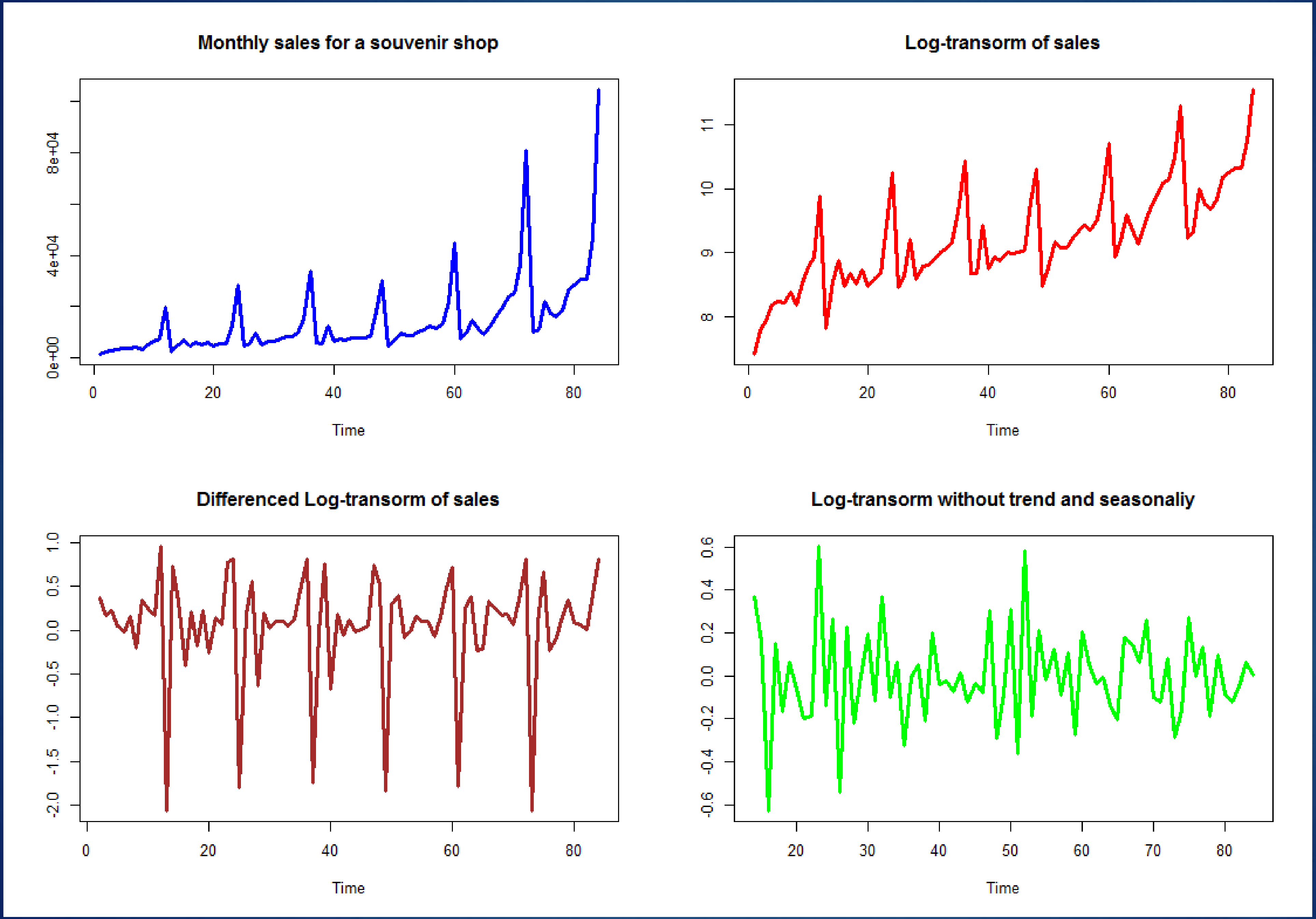

Monthly sales for a souvenir shop in Queensland, Australia. Jan 1987-Dec 1993
0e+00
2e+04
4e+04
6e+04
8e+04
1e+05
0
20
40
60
80
Time

Log-transform, non-seasonal and seasonal differencing

$$d = 1$$

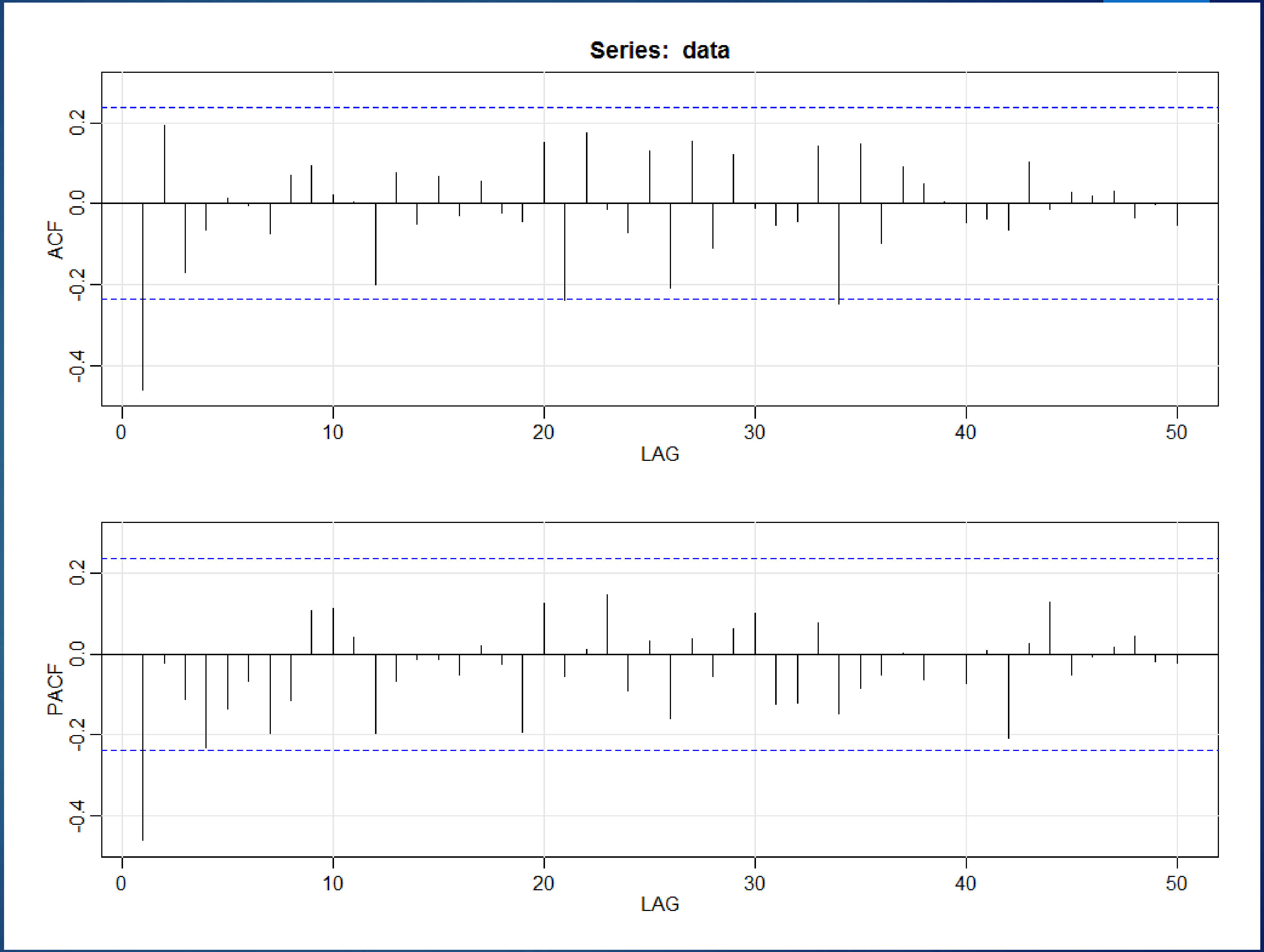
$$D = 1$$

$$diff(diff(\log()), 12)$$


Monthly sales for a souvenir shop
0e+00
4e+04
8e+04
0
20
40
60
80
Time
Log-transorm of sales
8
9
10
11
0
20
40
60
80
Time
Differenced Log-transorm of sales
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
0
20
40
60
80
Time
Log-transorm without trend and seasonaliy
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
20
30
40
50
60
70
80
Time

ACF

PACF


Series: data
ACF
0.2
0.0
-0.2
-0.4
0
10
20
30
40
50
LAG
PACF
0.2
0.0
-0.2
-0.4
0
10
20
30
40
50
LAG

Order specification

- ACF → $q = 0,1 \; ; Q = 0, 1,2,3$

- PACF → $p = 0,1; \mathrm{P} = 0, 1$

0 1 0 0 1 0 12 AIC= -11.60664 SSE= 3.432906 p-VALUE= 0.0001365566

0 1 0 0 1 1 12 AIC= -16.09179 SSE= 2.97756 p-VALUE= 3.149952e-05

0 1 0 0 1 2 12 AIC= -17.58234 SSE= 2.301963 p-VALUE= 0.0002456591

0 1 0 0 1 3 12 AIC= -16.41016 SSE= 2.35266 p-VALUE= 0.0003392283

0 1 0 1 1 0 12 AIC= -13.43083 SSE= 3.214065 p-VALUE= 4.083839e-05

0 1 0 1 1 1 12 AIC= -17.76362 SSE= 2.399746 p-VALUE= 0.0001916565

0 1 0 1 1 2 12 AIC= -15.99095 SSE= 2.349897 p-VALUE= 0.0002477782

0 1 0 1 1 3 12 AIC= -14.74777 SSE= 2.302026 p-VALUE= 0.0004504601

0 1 1 0 1 0 12 AIC= -27.78538 SSE= 2.643277 p-VALUE= 0.1742478

0 1 1 0 1 1 12 AIC= -34.54538 SSE= 2.233424 p-VALUE= 0.2730783

0 1 1 0 1 2 12 AIC= -33.6145 SSE= 2.109473 p-VALUE= 0.2830597

0 1 1 0 1 3 12 AIC= -32.19273 SSE= 1.87789 p-VALUE= 0.270042

0 1 1 1 1 0 12 AIC= -32.33192 SSE= 2.360507 p-VALUE= 0.2584529

0 1 1 1 1 1 12 AIC= -34.0881 SSE= 1.842013 p-VALUE= 0.2843225

0 1 1 1 1 2 12 AIC= -32.1017 SSE= 1.856342 p-VALUE= 0.28516

1 1 0 0 1 0 12 AIC= -27.07825 SSE= 2.6747 p-VALUE= 0.2297871

1 1 0 0 1 1 12 AIC= -34.98918 SSE= 2.209442 p-VALUE= 0.4633806

1 1 0 0 1 2 12 AIC= -33.38623 SSE= 2.159411 p-VALUE= 0.4515394

1 1 0 0 1 3 12 AIC= -31.54519 SSE= 2.121635 p-VALUE= 0.4390829

1 1 0 1 1 0 12 AIC= -32.64858 SSE= 2.340077 p-VALUE= 0.4022223

1 1 0 1 1 1 12 AIC= -33.48894 SSE= 2.125766 p-VALUE= 0.4442669

1 1 0 1 1 2 12 AIC= -31.52137 SSE= 2.093124 p-VALUE= 0.4463098

1 1 1 0 1 0 12 AIC= -26.17089 SSE= 2.624281 p-VALUE= 0.2507443

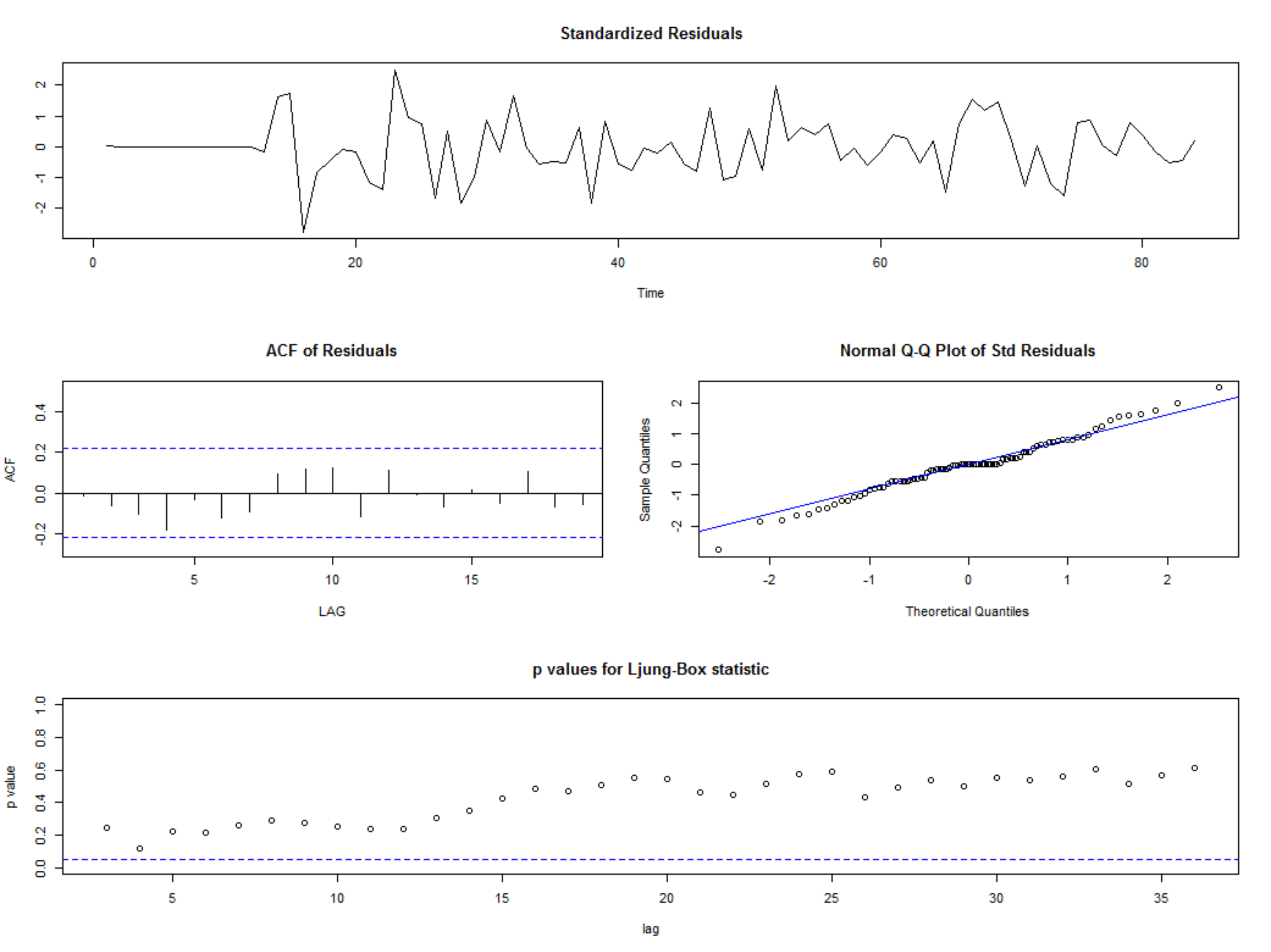
1 1 1 0 1 1 12 AIC= -33.30647 SSE= 2.201798 p-VALUE= 0.411014

1 1 1 0 1 2 12 AIC= -31.68924 SSE= 2.151774 p-VALUE= 0.3820814

1 1 1 1 1 0 12 AIC= -31.10127 SSE= 2.323818 p-VALUE= 0.3492746

1 1 1 1 1 1 12 AIC= -32.69913 **SSE= 1.824041** p-VALUE= 0.3092406

Residual analysis - SARIMA$(1,1,0,0,1,1)_{12}$



$SARIMA(1,1,0,0,1,1)_{12}$

	Estimate	SE	t.value	p.value
ar1	-0.5017	0.1013	-4.9531	0.0000
sma1	-0.5107	0.1543	-3.3098	0.0014

Model – SARIMA$(1,1,0,0,1,1)_{12}$

$$X_t = Sales\ at\ a\ souvenir\ shop$$

$$Y_t = \log(X_t)$$

$$(1-\phi B)(1-B)(1-B^{12})Y_t = (1+\Theta B^{12})Z_t$$

$$Y_t = (1+\phi)Y_{t-1} - \phi Y_{t-2} - (1+\phi)Y_{t-13} + \phi Y_{t-14} + Z_t + \Theta\, Z_{t-12}$$

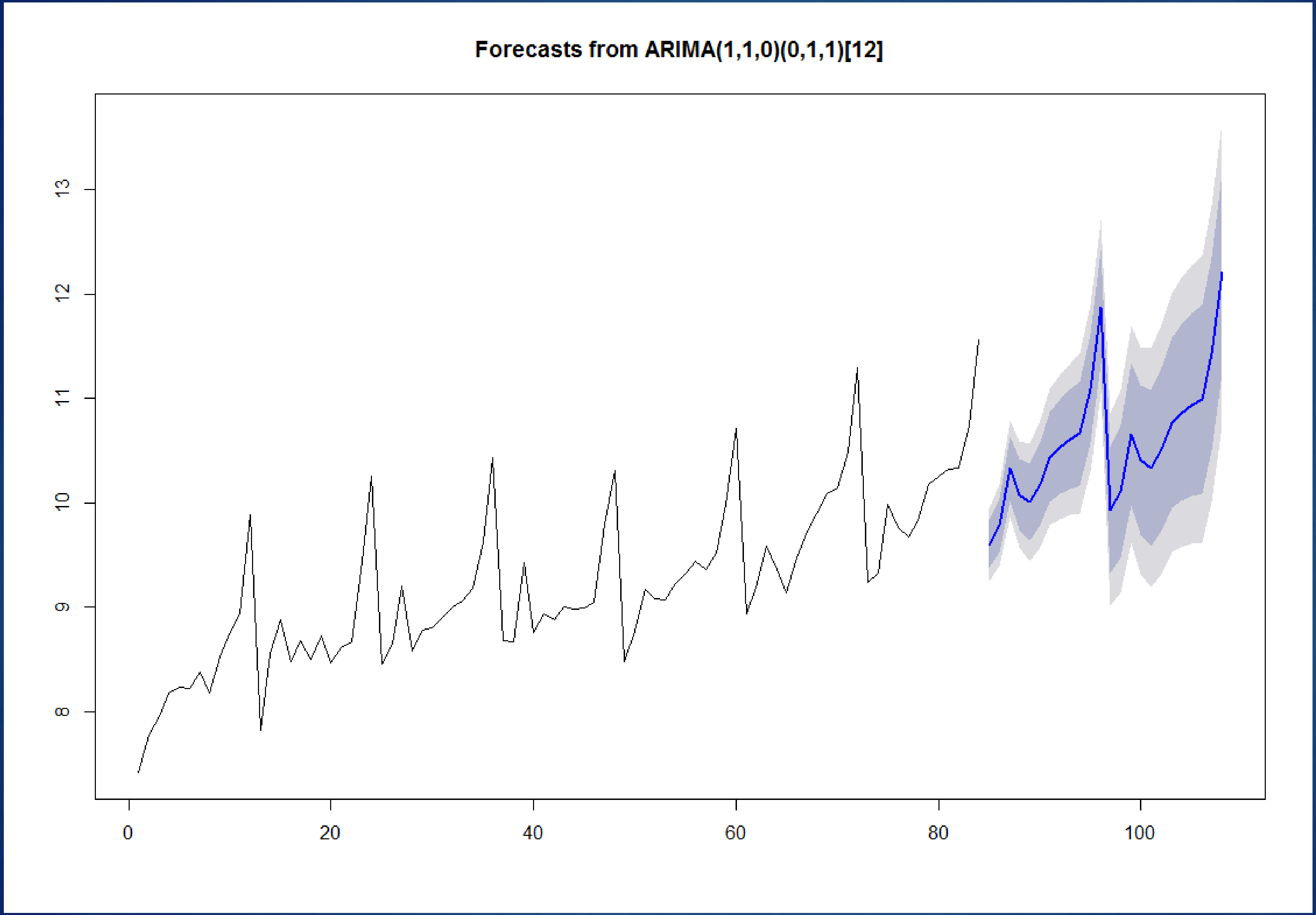
$$\hat{\phi} = -0.5017, \qquad \hat{\Theta} = -0.5107$$

Model – cont.

$$Y_t = 0.4983\, Y_{t-1} + 0.5017\, Y_{t-2} - 0.4983\, Y_{t-13} - 0.5017\, Y_{t-14} + Z_t - 0.5107\, Z_{t-12}$$

where

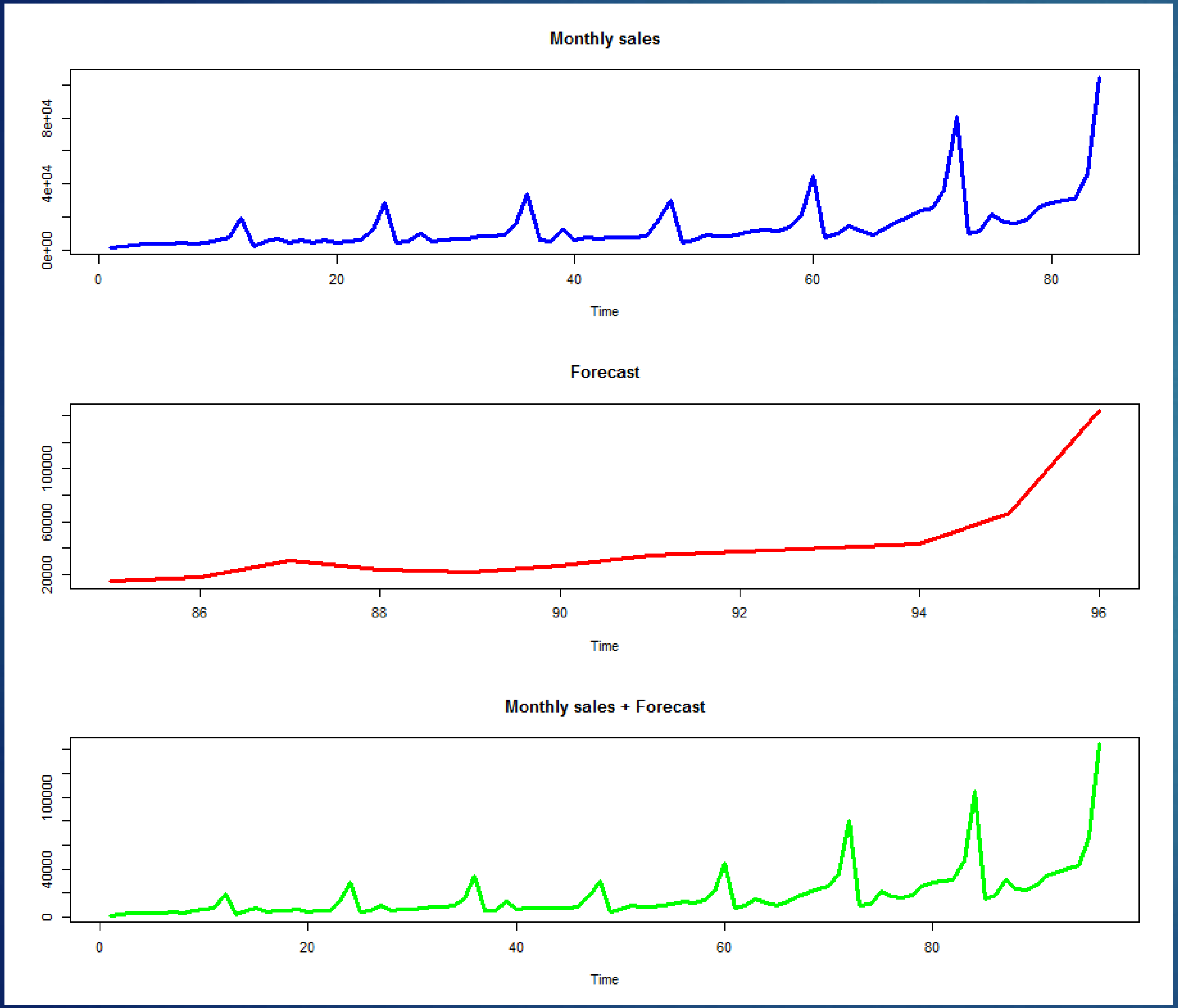
$$Z_t \sim Normal\ (0, 0.0311)$$

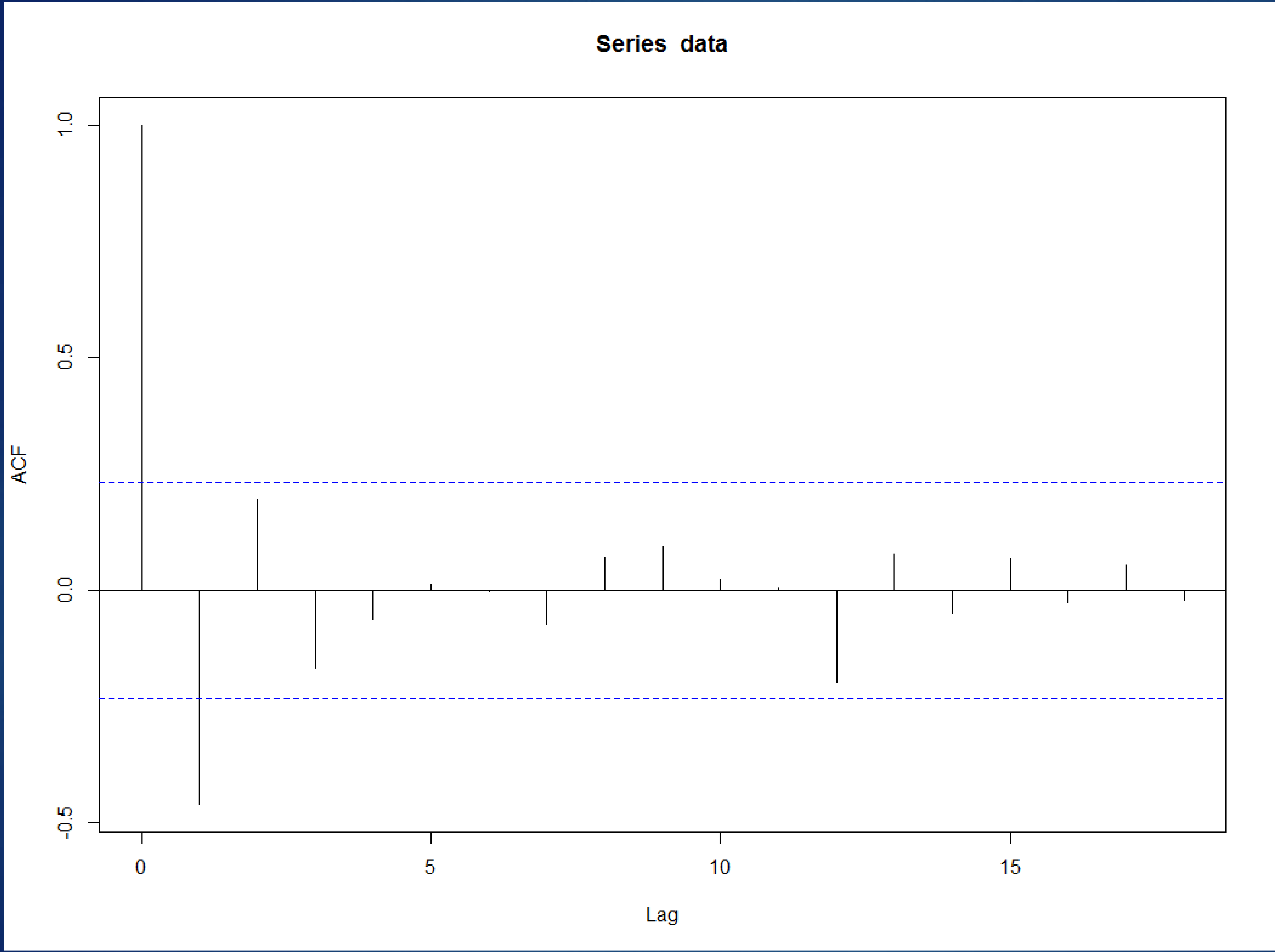

Forecasts from ARIMA(1,1,0)(0,1,1)[12]
8
9
10
11
12
13
0
20
40
60
80
100

forecast(model)

	Pt. for.	Lo 80	Hi 80	Lo 95	Hi 95
85	9.600019	9.373968	9.826071	9.254303	9.945736
86	9.786505	9.533944	10.039066	9.400246	10.172764
87	10.329605	10.025423	10.633786	9.864399	10.794810
88	10.081973	9.746705	10.417240	9.569225	10.594720
89	10.008096	9.638604	10.377587	9.443007	10.573184
90	10.181170	9.783094	10.579245	9.572365	10.789974
91	10.439372	10.013362	10.865383	9.787845	11.090900
92	10.534857	10.083237	10.986477	9.844164	11.225551
93	10.613026	10.136886	11.089165	9.884833	11.341218
94	10.664526	10.165207	11.163846	9.900883	11.428170
95	11.096784	10.575248	11.618321	10.299163	11.894406
96	11.877167	11.334355	12.419979	11.047007	12.707326

Data + Forecast

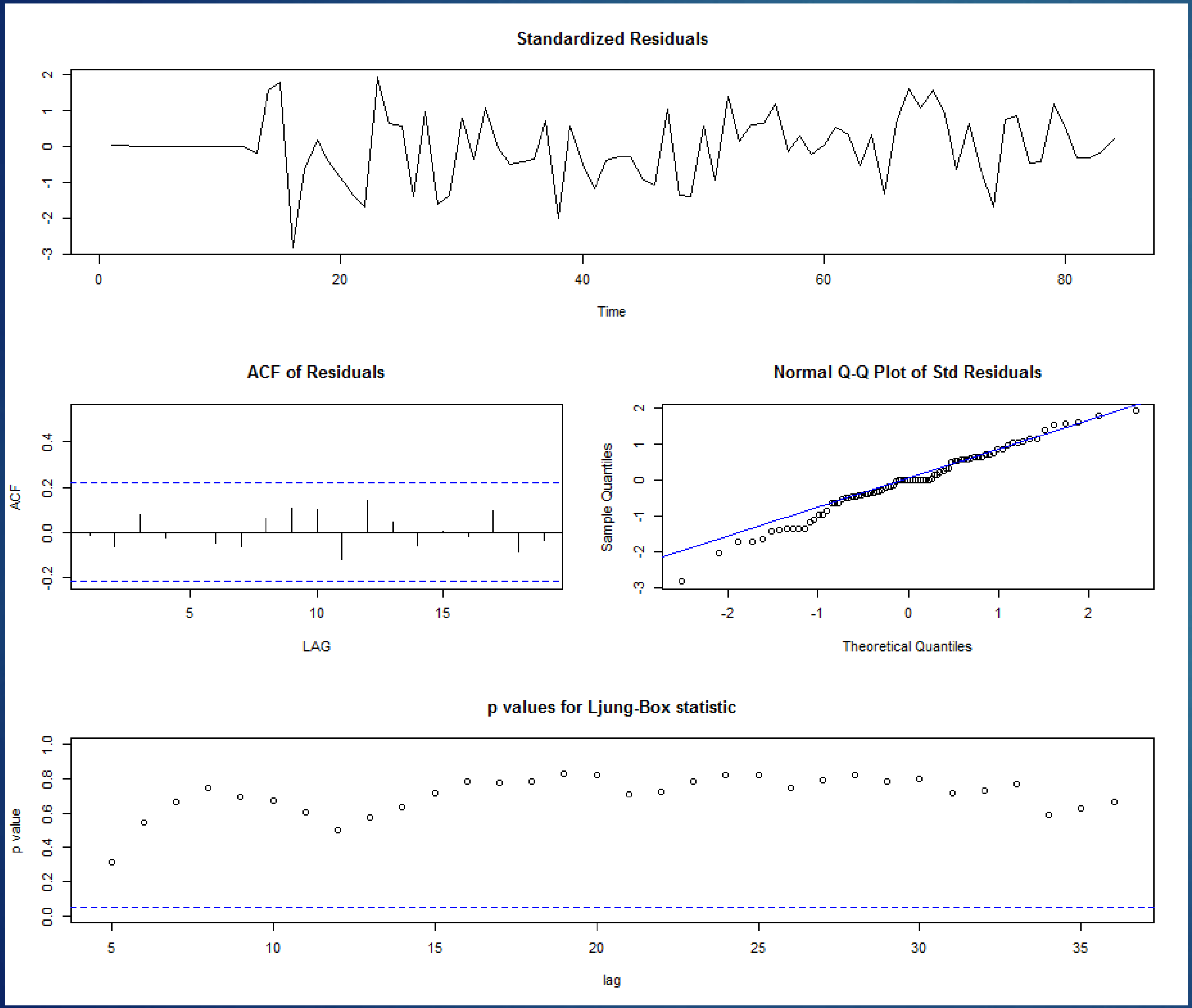



Series data
ACF
1.0
0.5
0.0
-0.5
0
5
10
15
Lag

Model comparison

	SARIMA$(1,1,0,0,1,1)_{12}$	**SARIMA**$(0,1,3,0,1,1)_{12}$
AIC	-34.99	-37.56
SSE	2.21	1.99
p-value	0.46	0.97

Residual analysis - SARIMA$(0,1,3,0,1,1)_{12}$



What We've Learned

- Fit SARIMA models to dataset about sales at a souvenir shop from TSDL
- Forecast future values of examined time series