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$(p, d, q, P, D, Q)_s$

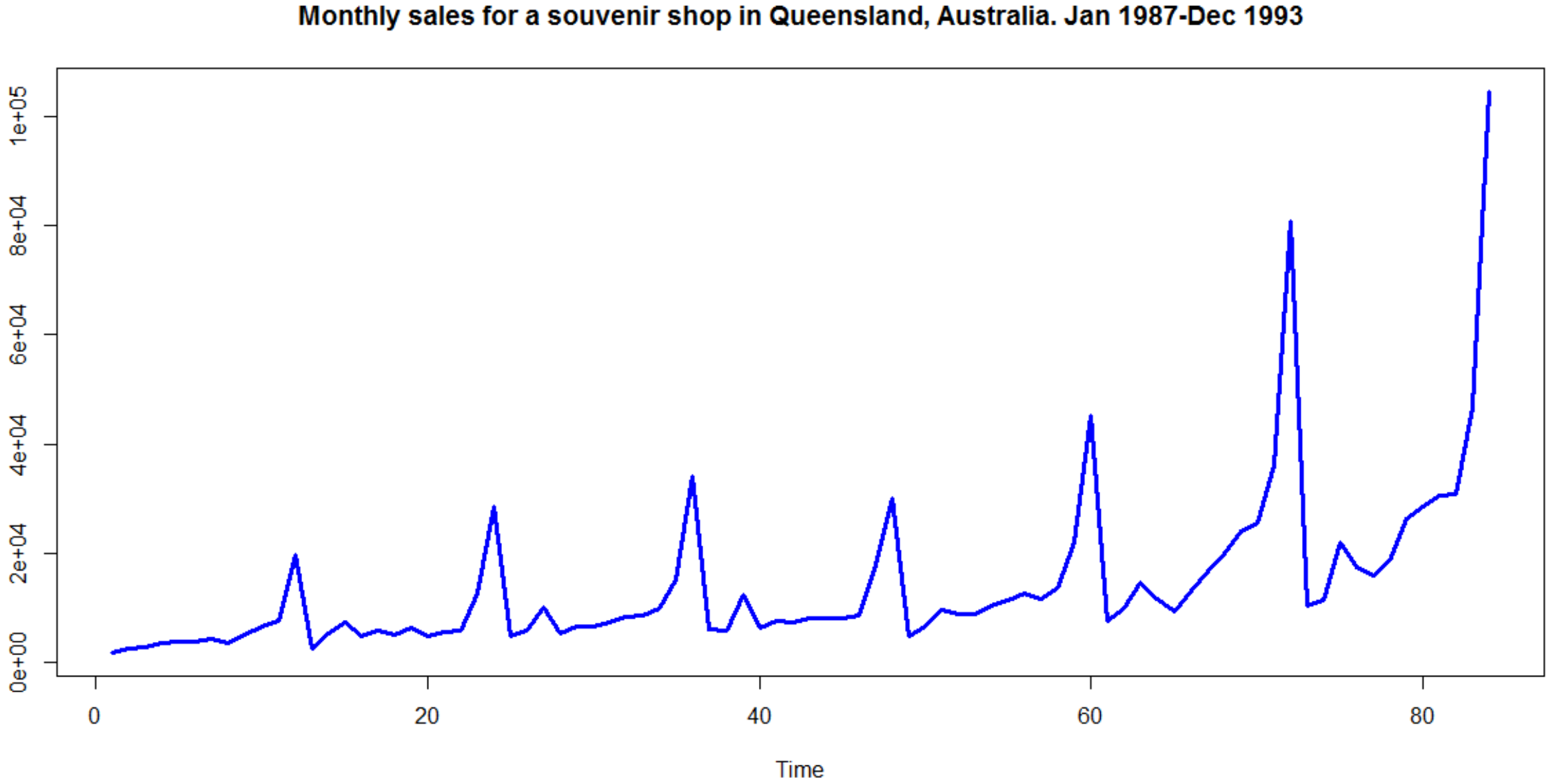
$$p + d + q + P + D + Q \leq 6$$

Time Series Data Library

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

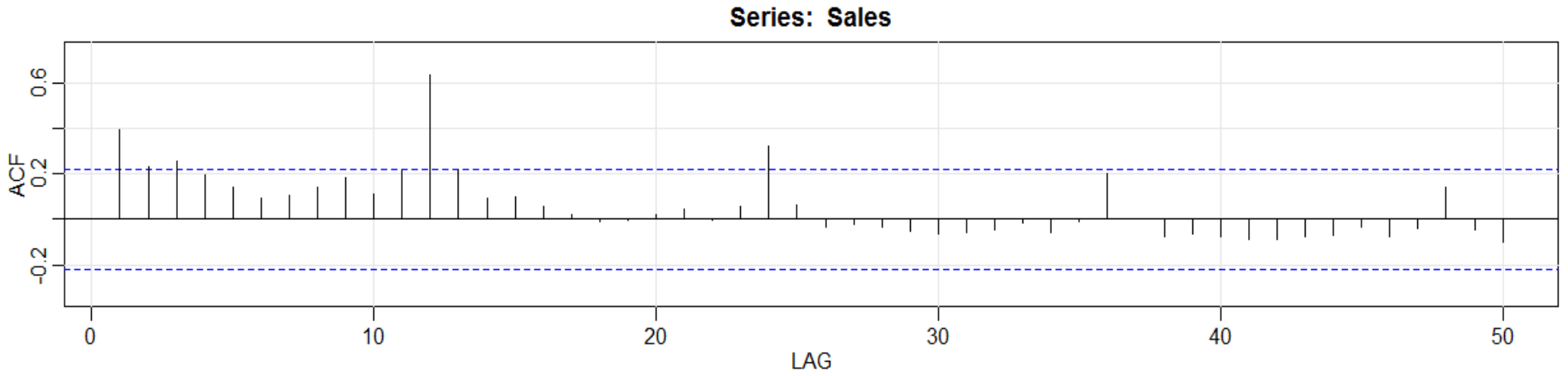
Monthly sales for a souvenir shop: Sales

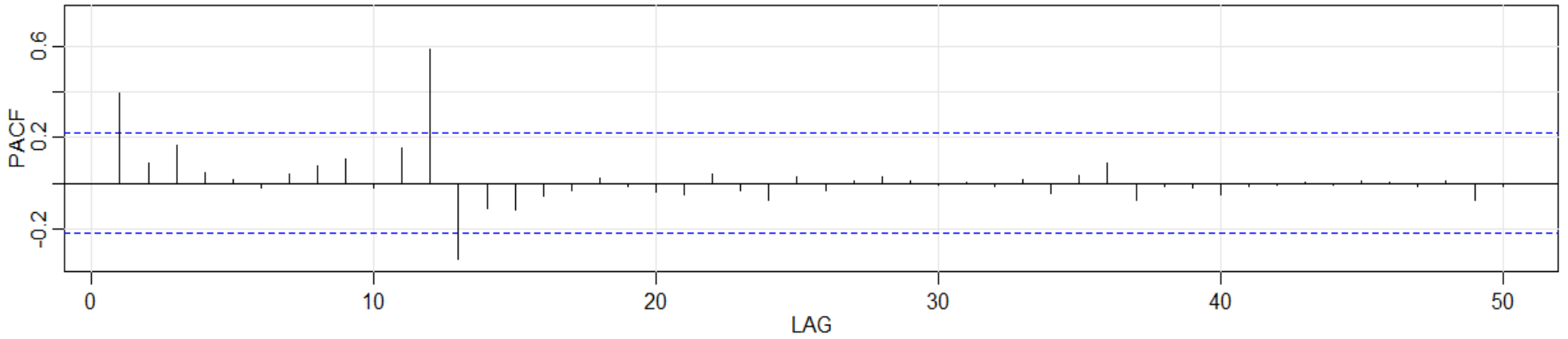
- https://datamarket.com/data/set/22mh/monthly-sales-for-a-souvenir-shop-on-the-wharf-at-a-beach-resort-town-in-queensland-australia-jan-1987-dec-1993#!ds=22mh&display=line
- Sales for a souvenir shop in Queensland, Australia
- January 1987 – December 1993
- Sales, Source: Makridakis, Wheelwright and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

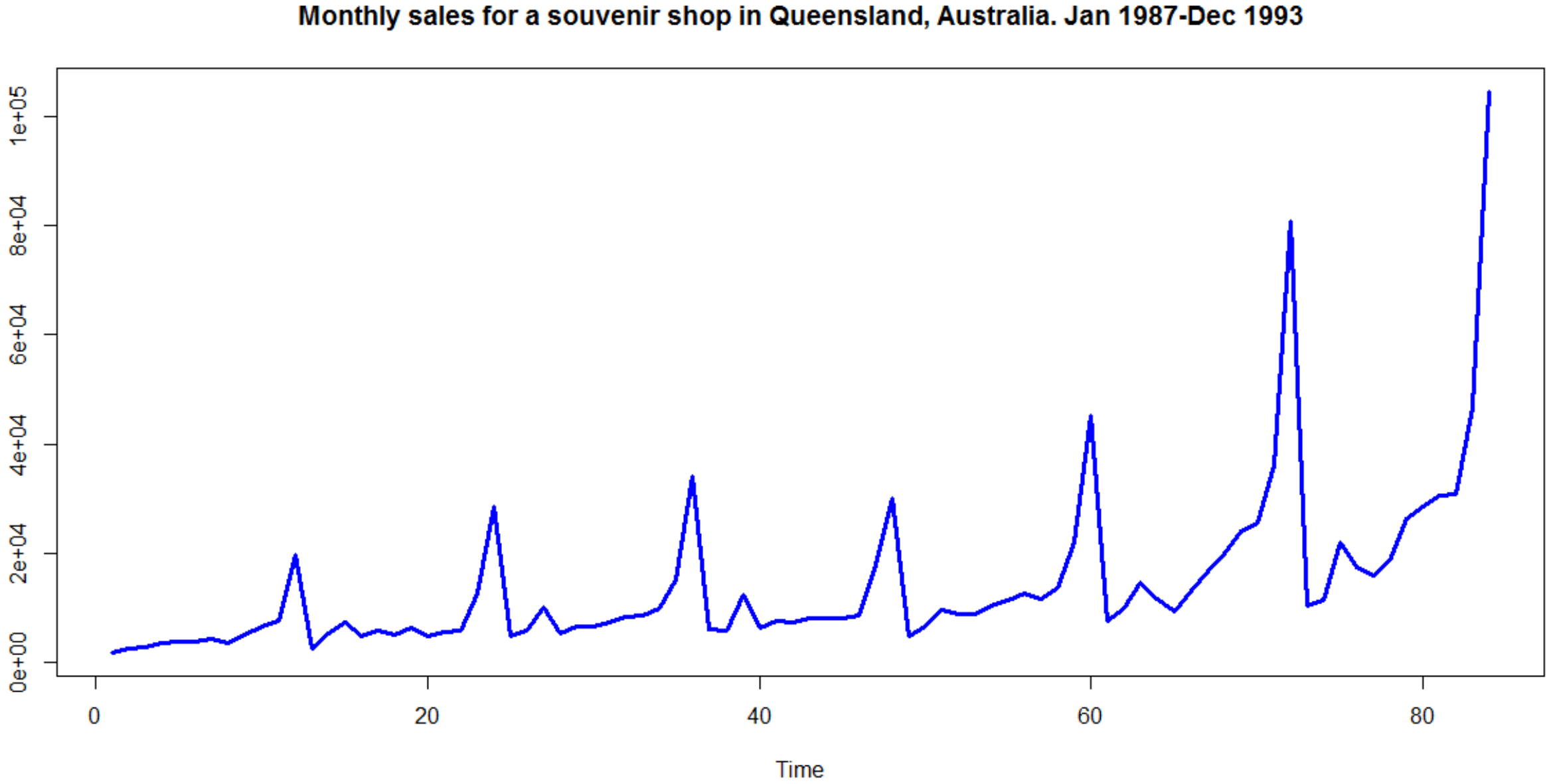

Monthly sales for a souvenir shop in Queensland, Australia. Jan 1987-Dec 1993
0e+00
2e+04
4e+04
6e+04
8e+04
1e+05
0
20
40
60
80
Time

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

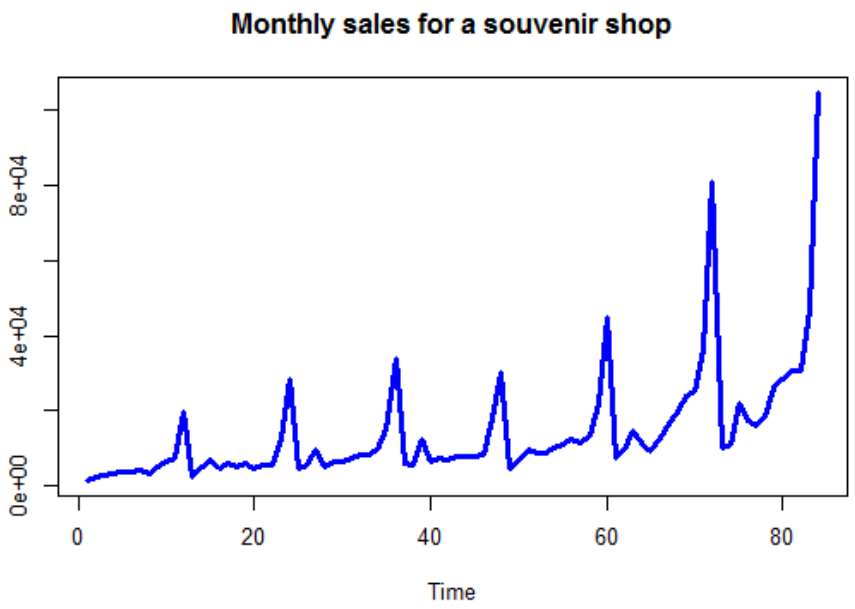

Monthly sales for a souvenir shop in Queensland, Australia. Jan 1987-Dec 1993
0e+00
2e+04
4e+04
6e+04
8e+04
1e+05
0
20
40
60
80
Time

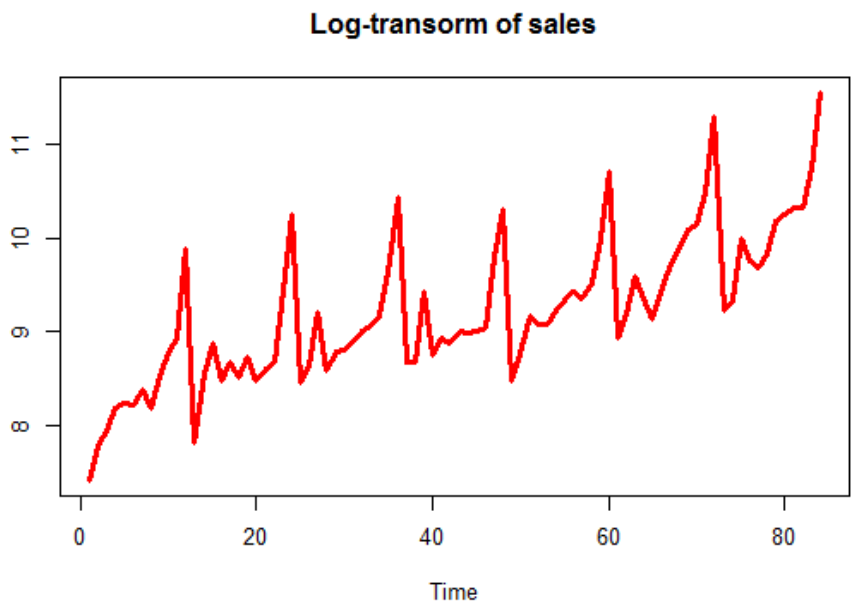
Log-transform, non-seasonal and seasonal differencing

$$d = 1$$

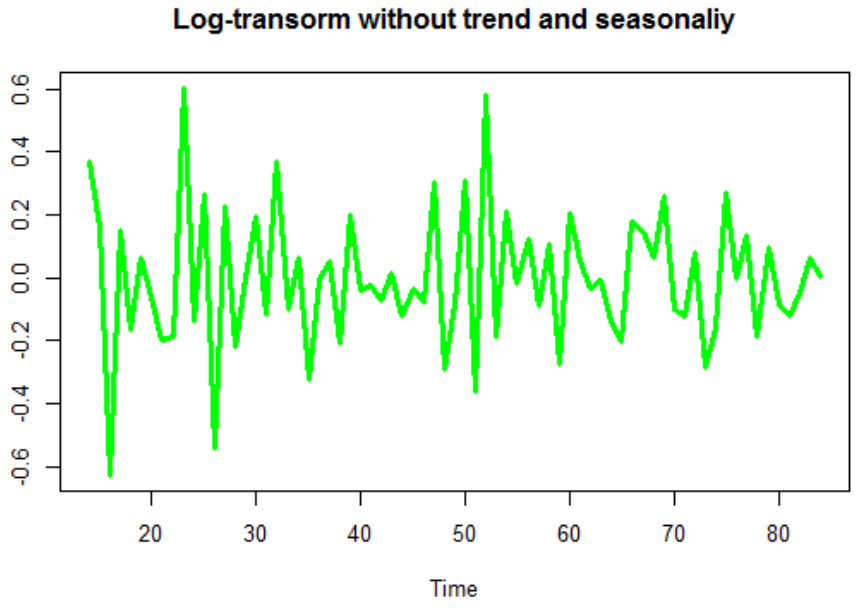
$$D = 1$$

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


Monthly sales for a souvenir shop
0e+00
4e+04
8e+04
0
20
40
60
80
Time

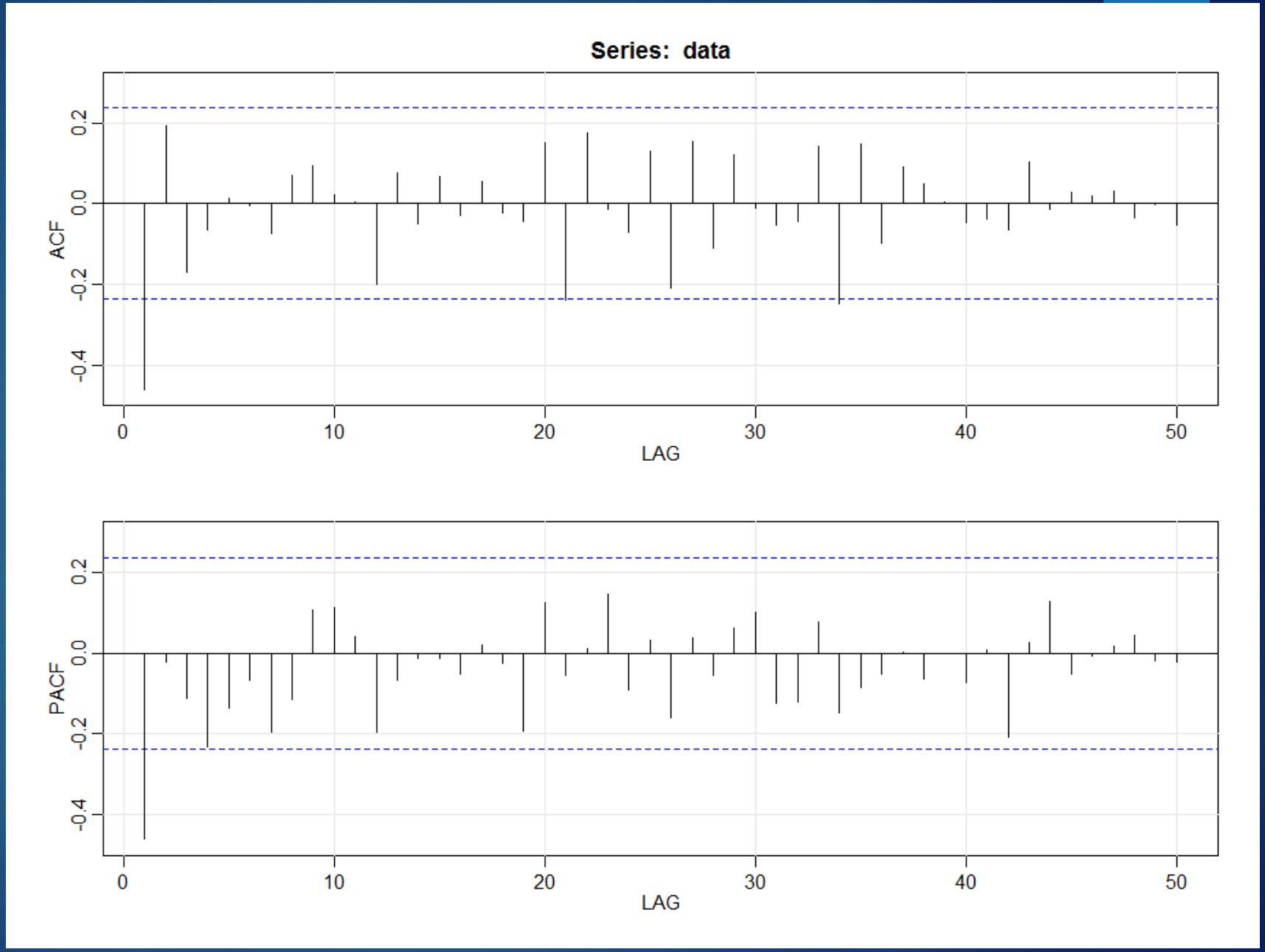

Log-transorm of sales
8
9
10
11
0
20
40
60
80
Time


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

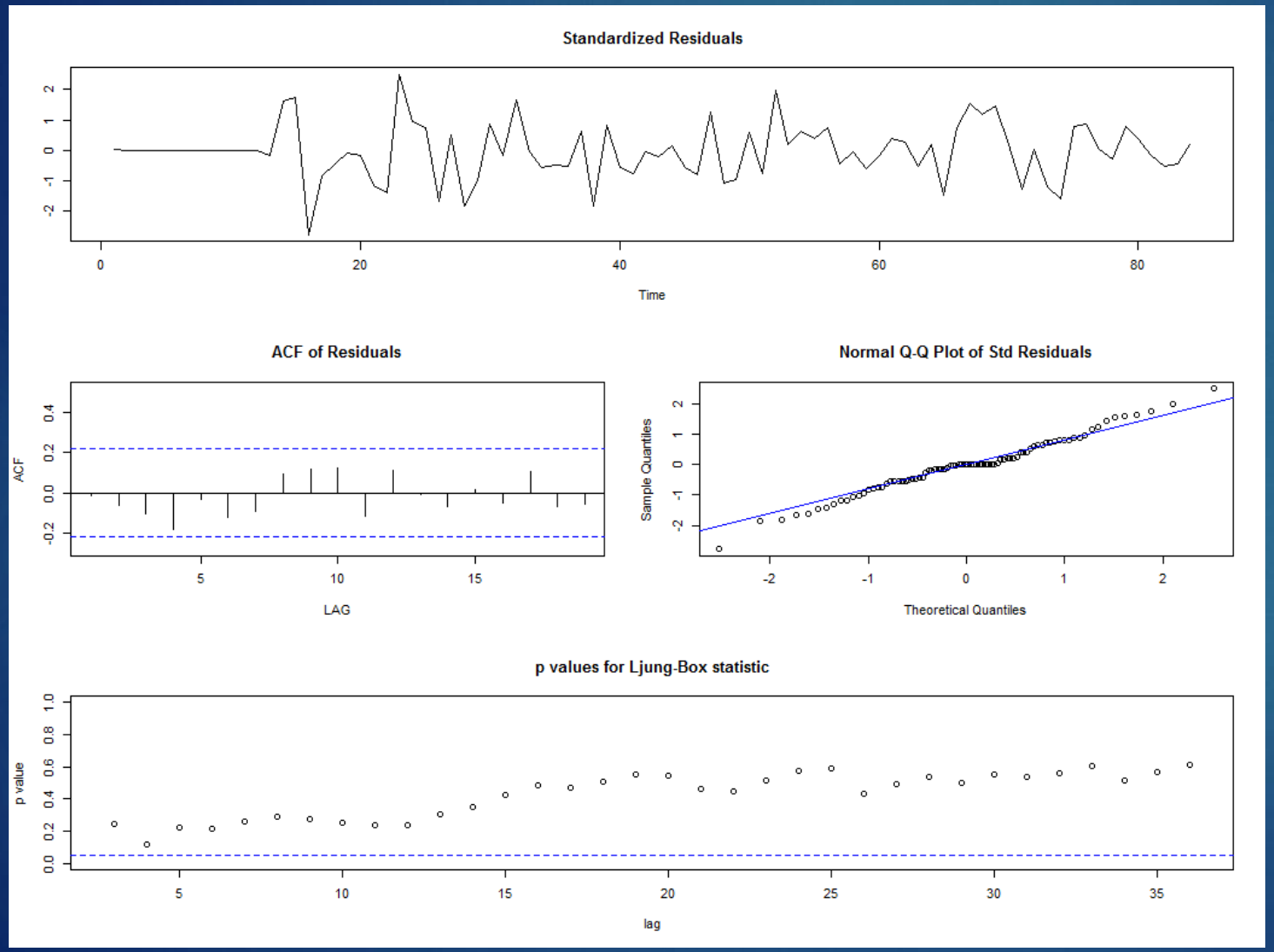
1 1 1 0 1 1 12 AIC= -33.30647 SSE= 2.201798 p-VALUE= 0.411014

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 – $SARIMA(1,1,0,0,1,1)_{12}$

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

$$Y_t = \log(X_t)$$

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

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

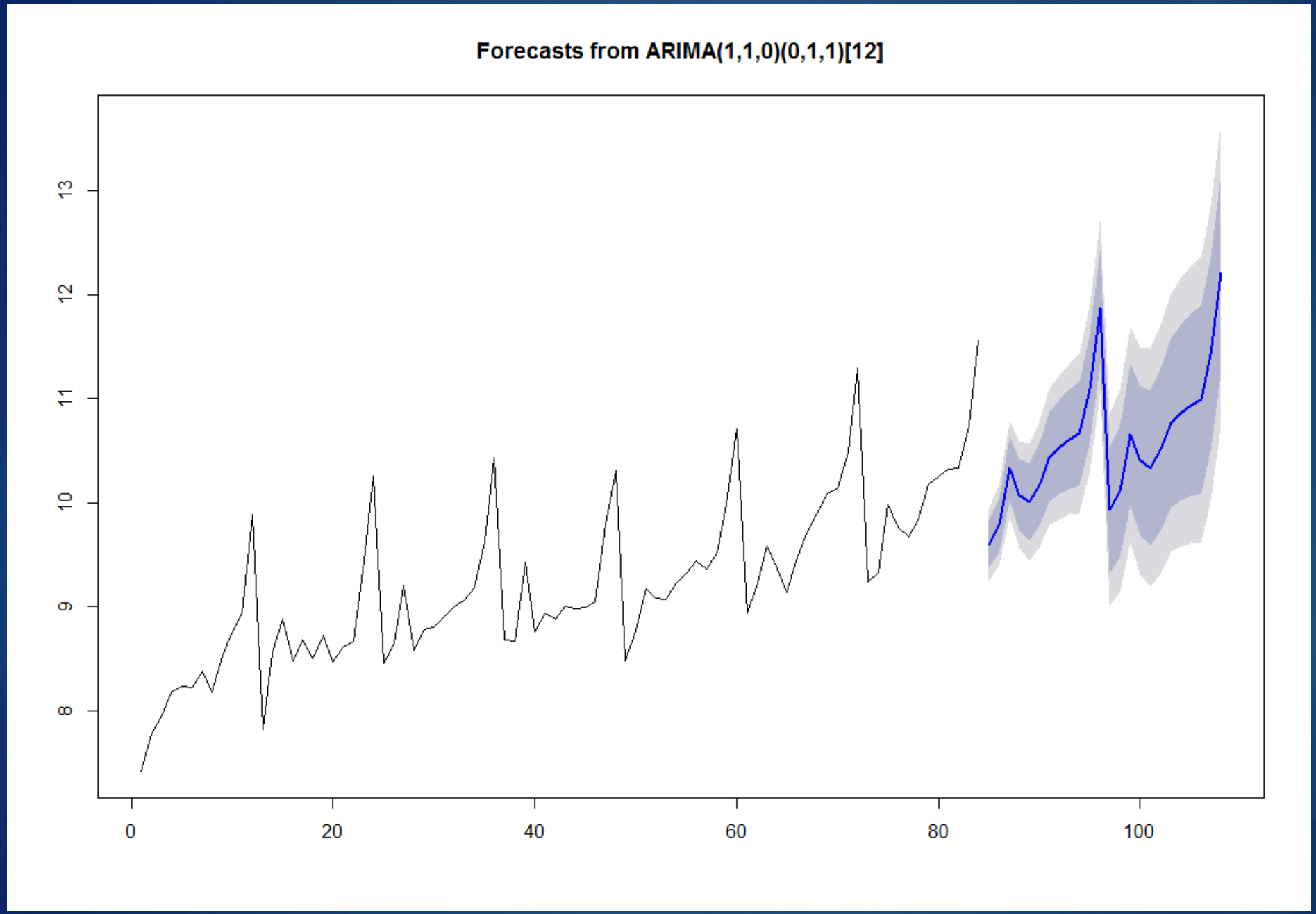
$$\hat{\phi} = -0.5017, \qquad \hat{\Theta} = -0.5107$$

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

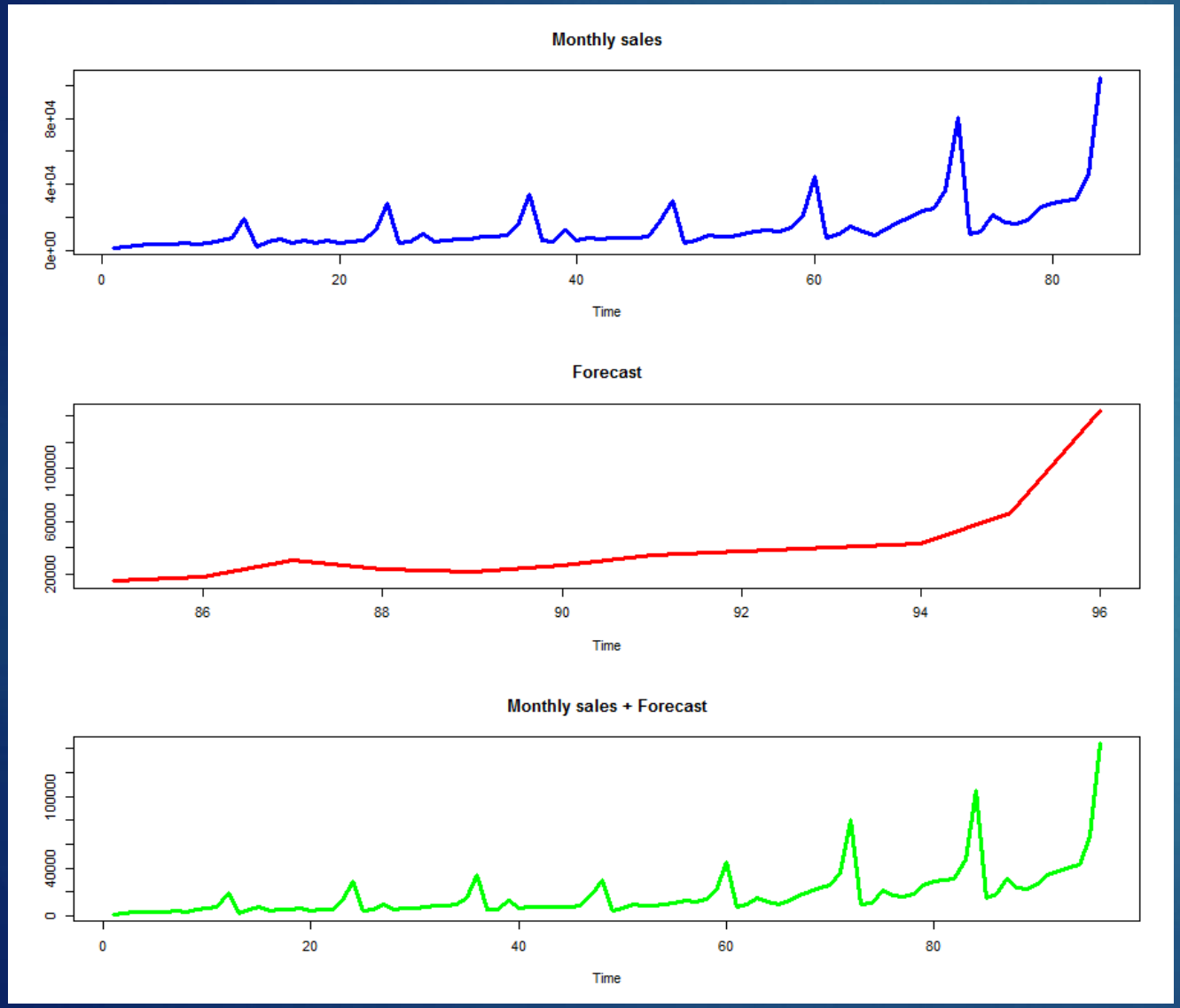
$$Z_t \sim Normal\ (0, 0.0311)$$

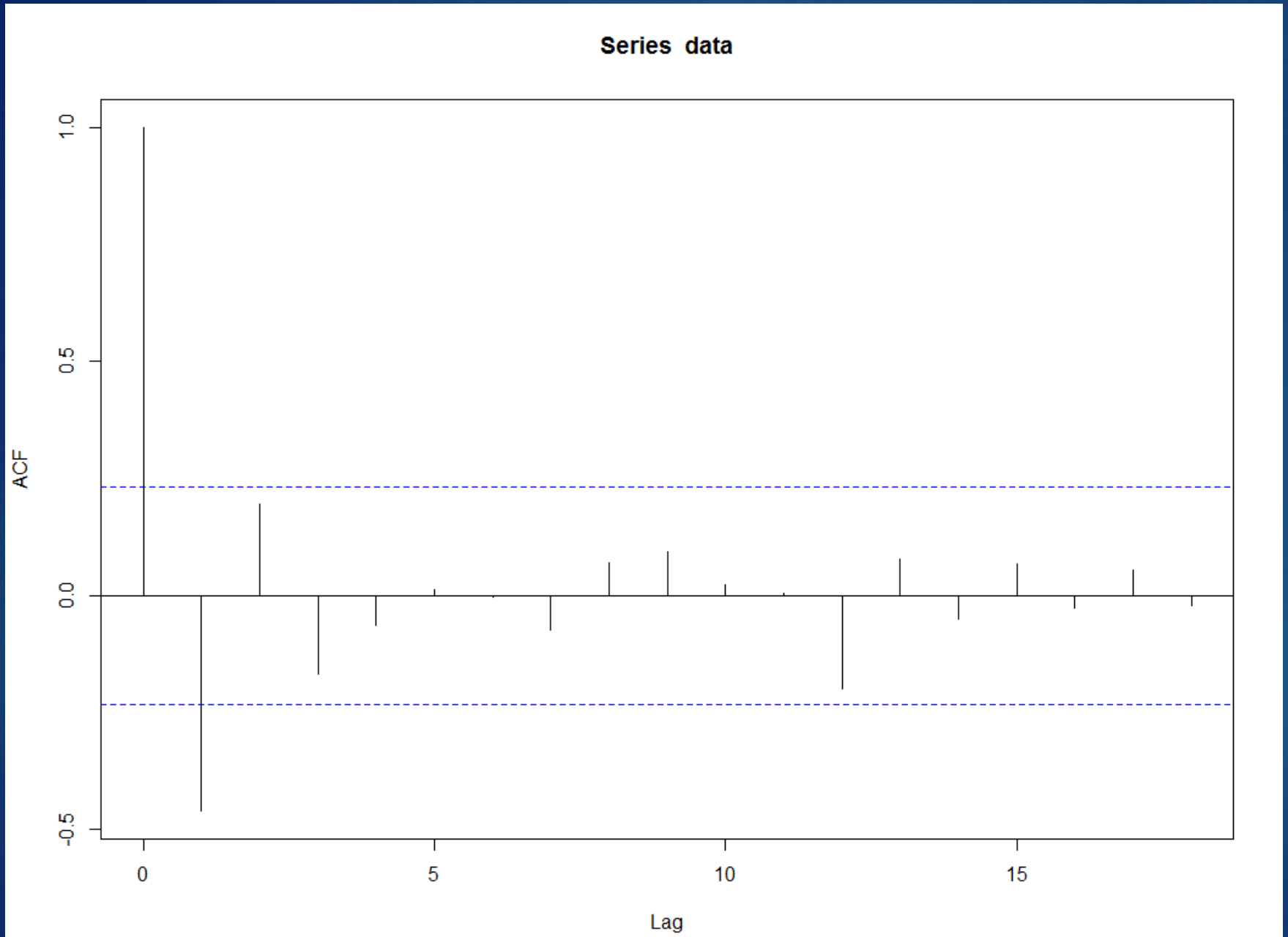

Forecasts from ARIMA(1,1,0)(0,1,1)[12]
8
9
10
11
12
13
0
20
40
60
80
100

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

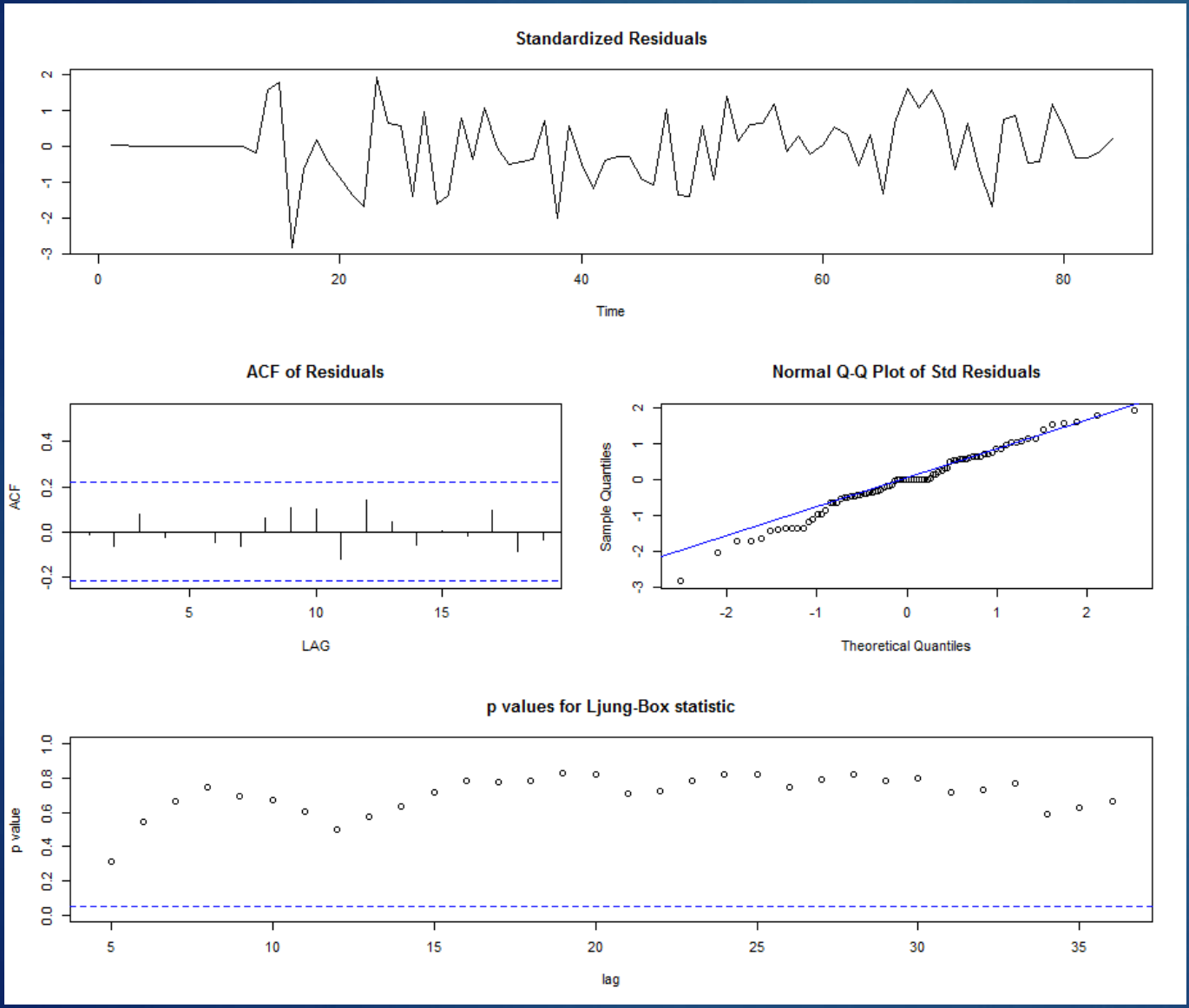



Series data
ACF
1.0
0.5
0.0
-0.5
0
5
10
15
Lag

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