TIME SERIES

Overview

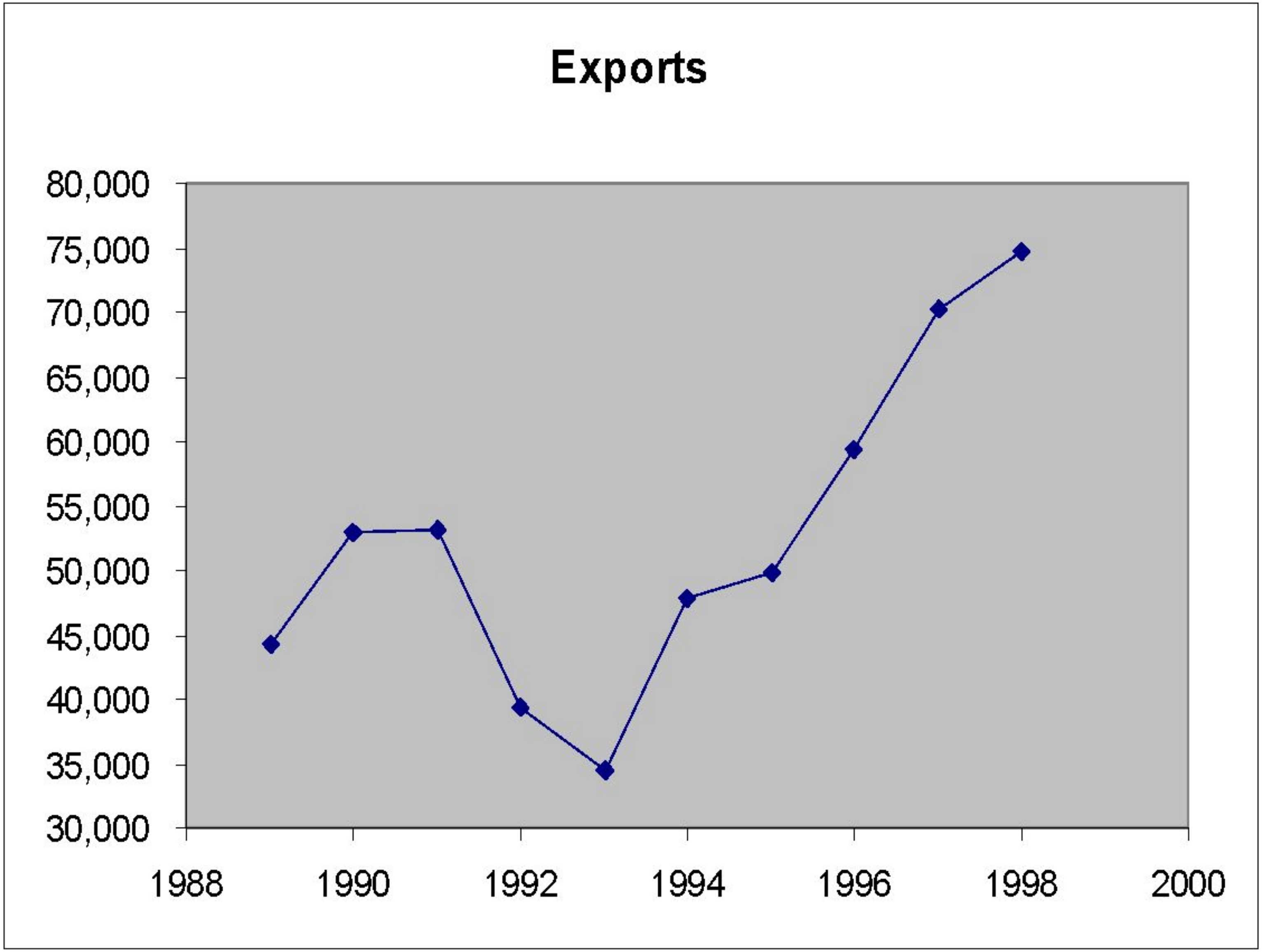
1. Time Series

2. Time Series Analysis
   - 2.1 Decomposition of time series
   - 2.2 Decomposition of Additive time series
   - 2.3 Decomposition of Multiplicative time series

3. Time Series Forecasting
   - 3.1 Preliminary analysis.
   - 3.2 Choosing and fitting models.
   - 3.3 Using and evaluating a forecasting model.

1. Time Series

- **Time Series** is a series of values of a quantity (variable) obtained at successive times, often with equal time intervals between them.



Exports
80,000
75,000
70,000
65,000
60,000
55,000
50,000
45,000
40,000
35,000
30,000
1988
1990
1992
1994
1996
1998
2000

WHY TIME SERIES ANALYSIS ?

- **Regression** $Y = a + b X$

- X Independent variable

- Y Dependent variable

Application:

1. Forecasting Inflation
2. Forecasting unemployment rate
3. Forecasting currency exchange rate
4. Forecasting gold or silver rate
5. Forecasting GDP

- **Time Series** is a series of values of a quantity (variable) obtained at successive times, often with equal time intervals between them.

- **Time series *analysis*** comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.

- **Time series *forecasting*** is the use of a model to predict future values based on previously observed values.

2. Time Series Analysis

COMPONENTS OF A TIME SERIES

The pattern in a time series is sometimes classified into four components.

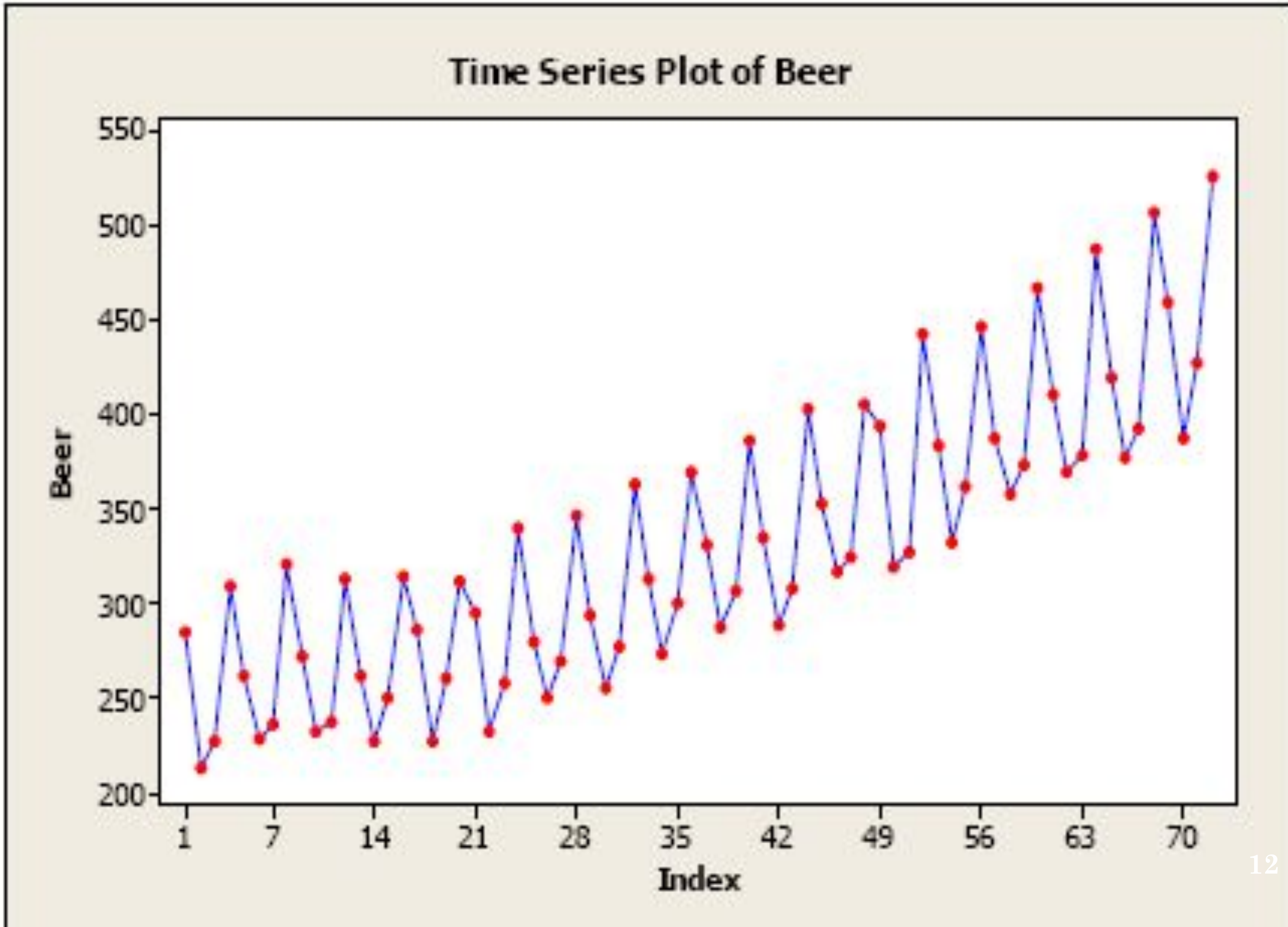
1. Trend
2. Seasonal
3. Cyclical
4. Random.

1. Trend component

- A trend exists when there is a long-term increase or decrease in the data.

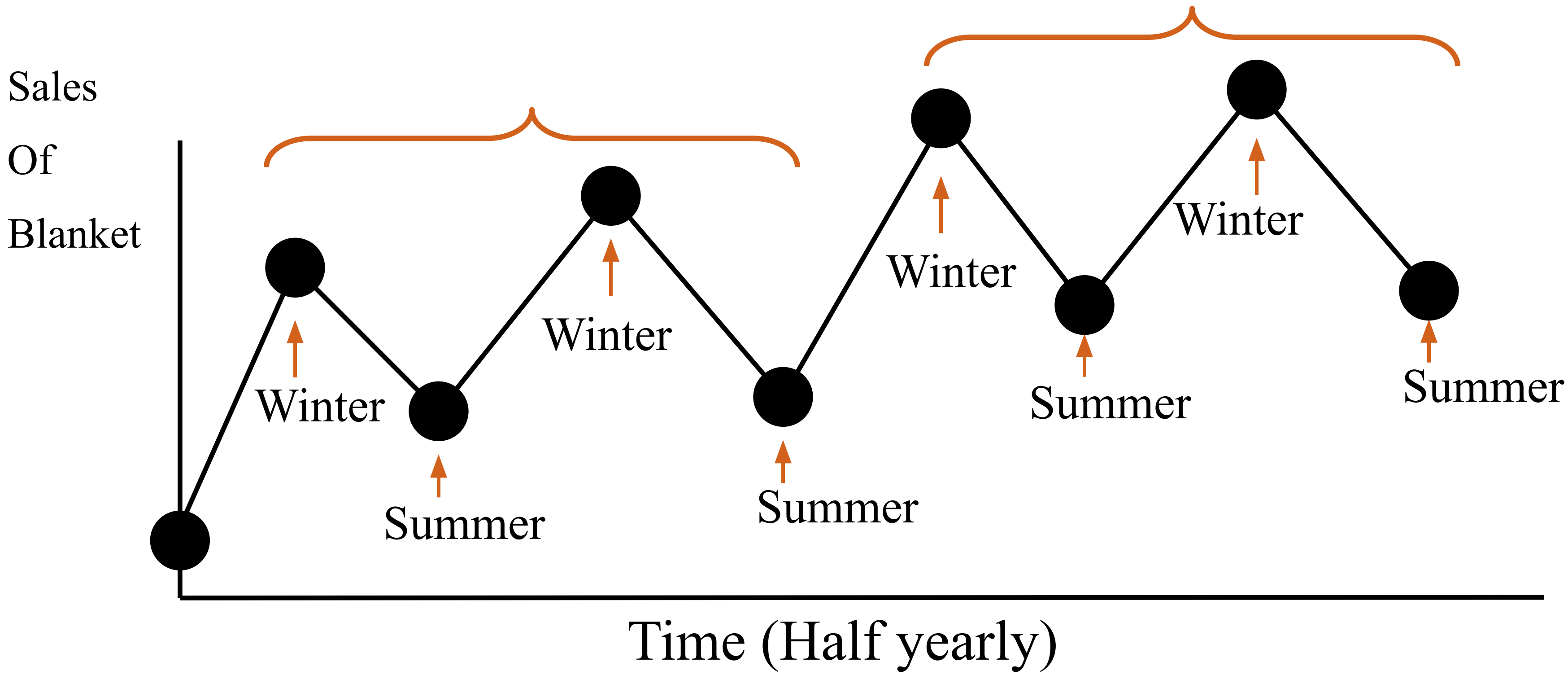
- The form of the trend pattern may be linear or non-linear

- Trend can change direction e. g. from an increasing trend to a decreasing trend.


Time Series Plot of Beer
550
500
450
400
350
300
250
200
Beer
1
7
14
21
28
35
42
49
56
63
70
Index

2. Seasonal Component:-

- Regular, relatively short-term repetitive up-and-down fluctuations of the values of variable.
- A seasonal pattern exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week).
- Seasonality is always of a fixed and known period.


Sales
Of
Blanket
Winter
Summer
Winter
Summer
Winter
Summer
Winter
Summer
Time (Half yearly)

3. Cyclical component

- A cyclic pattern exists when data exhibit rises and falls that are not of fixed period.
- The duration of these fluctuations is usually of at least 2 years.
- If the fluctuations are not of fixed period then they are cyclic.
- If the period is unchanging and associated with some aspect of the calendar, then the pattern is seasonal.
- Cyclical component is usually **not present** in the typical time series analysis.

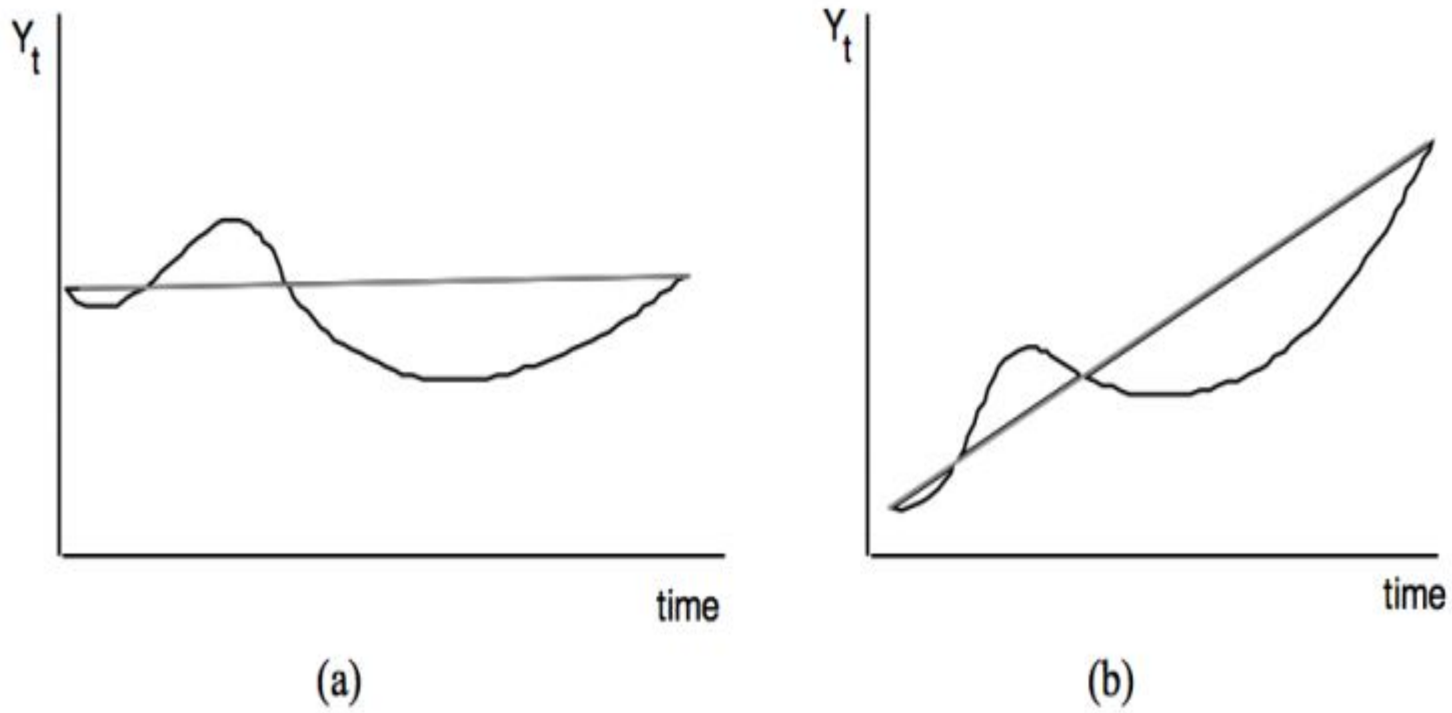
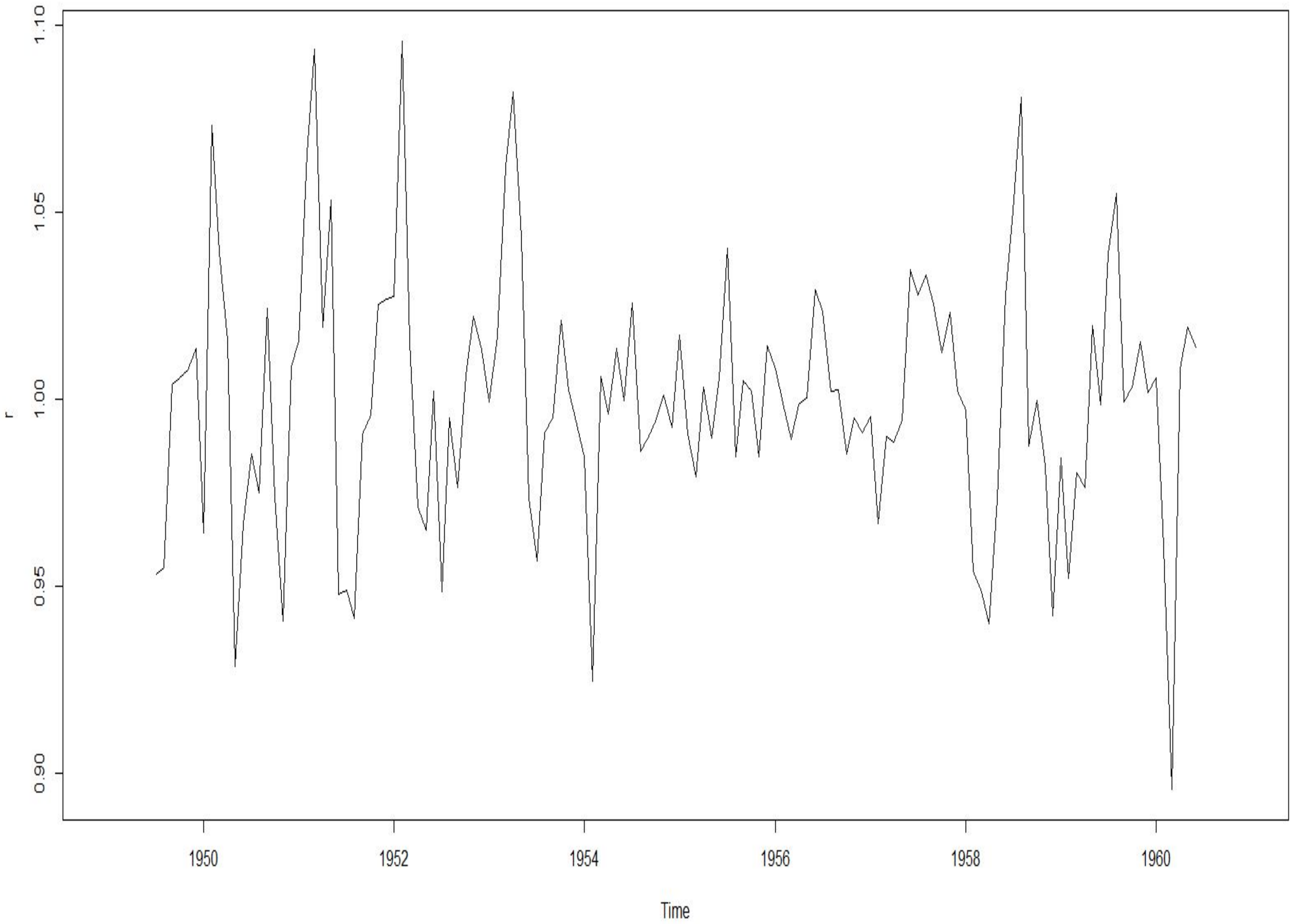


Figure 3: Cyclical component (imposed on the underlying trend).

4. Random / error component

- A random increase or decrease of values of the variable for a specific time period.
- This component is unpredictable.


1.10
1.05
1.00
0.95
0.90
r
1950
1952
1954
1956
1958
1960
Time

MODELS OF TIME SERIES

(a) Additive Model
(b) Multiplicative Model

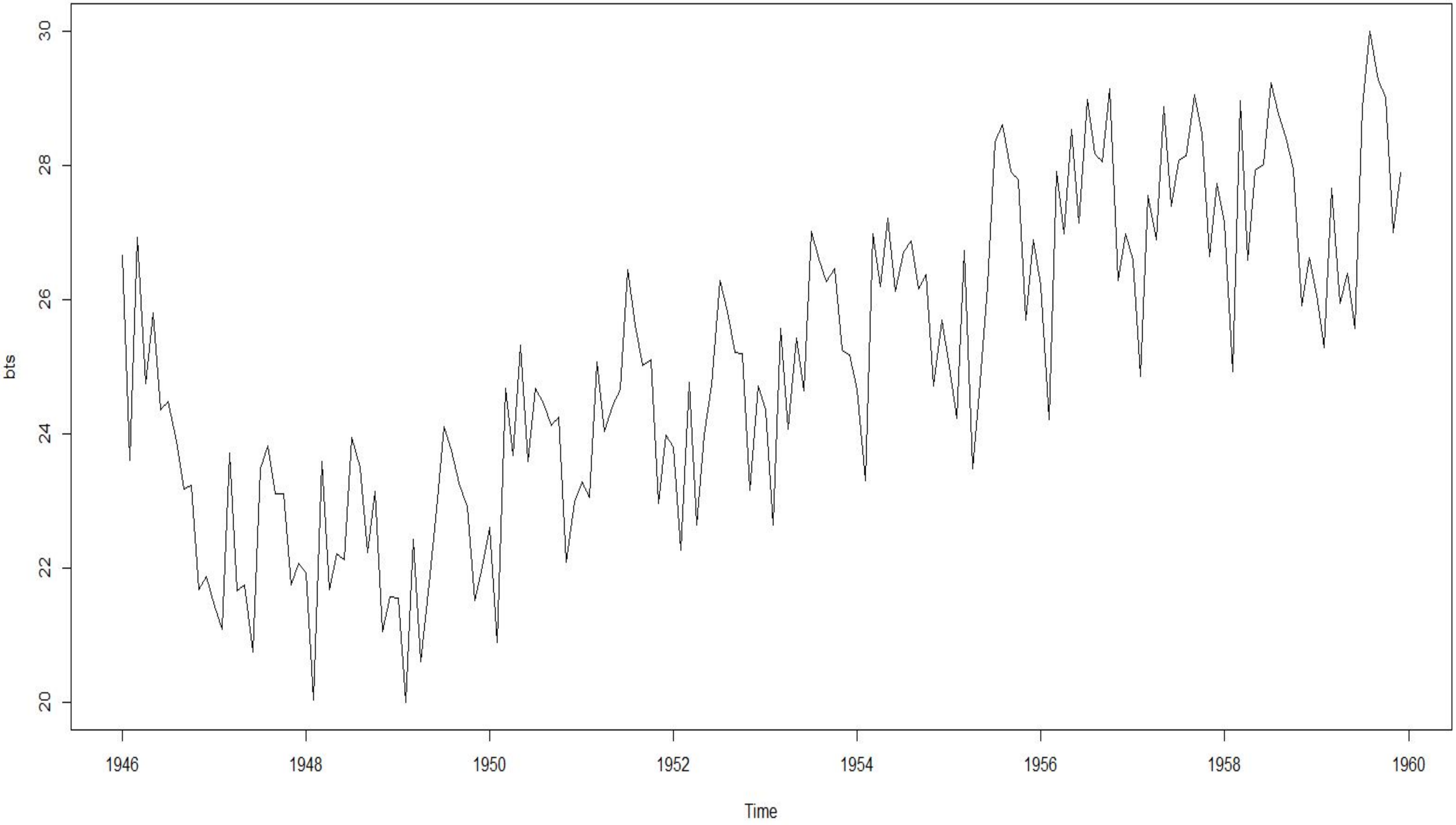
1. **Additive Model**:-

The additive model expresses value of Y at time t as the sum of the trend, cyclical, seasonal, and error components.

Additive Model of Time Series Data

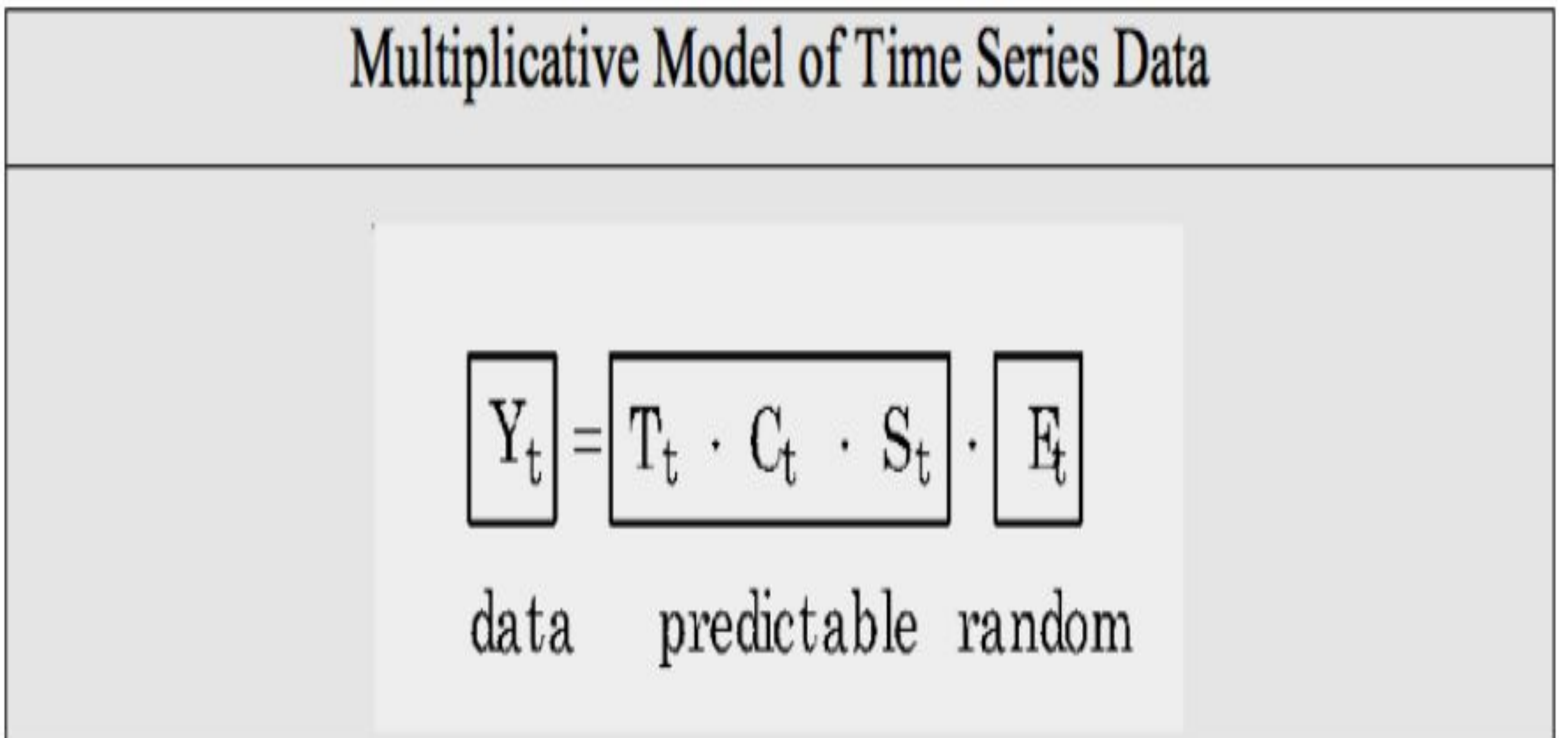
$$\underbrace{\boxed{Y_t}}_{\text{data}} = \underbrace{\boxed{T_t + C_t + S_t}}_{\text{predictable}} + \underbrace{\boxed{E_t}}_{\text{random}}$$

Example of Additive Model

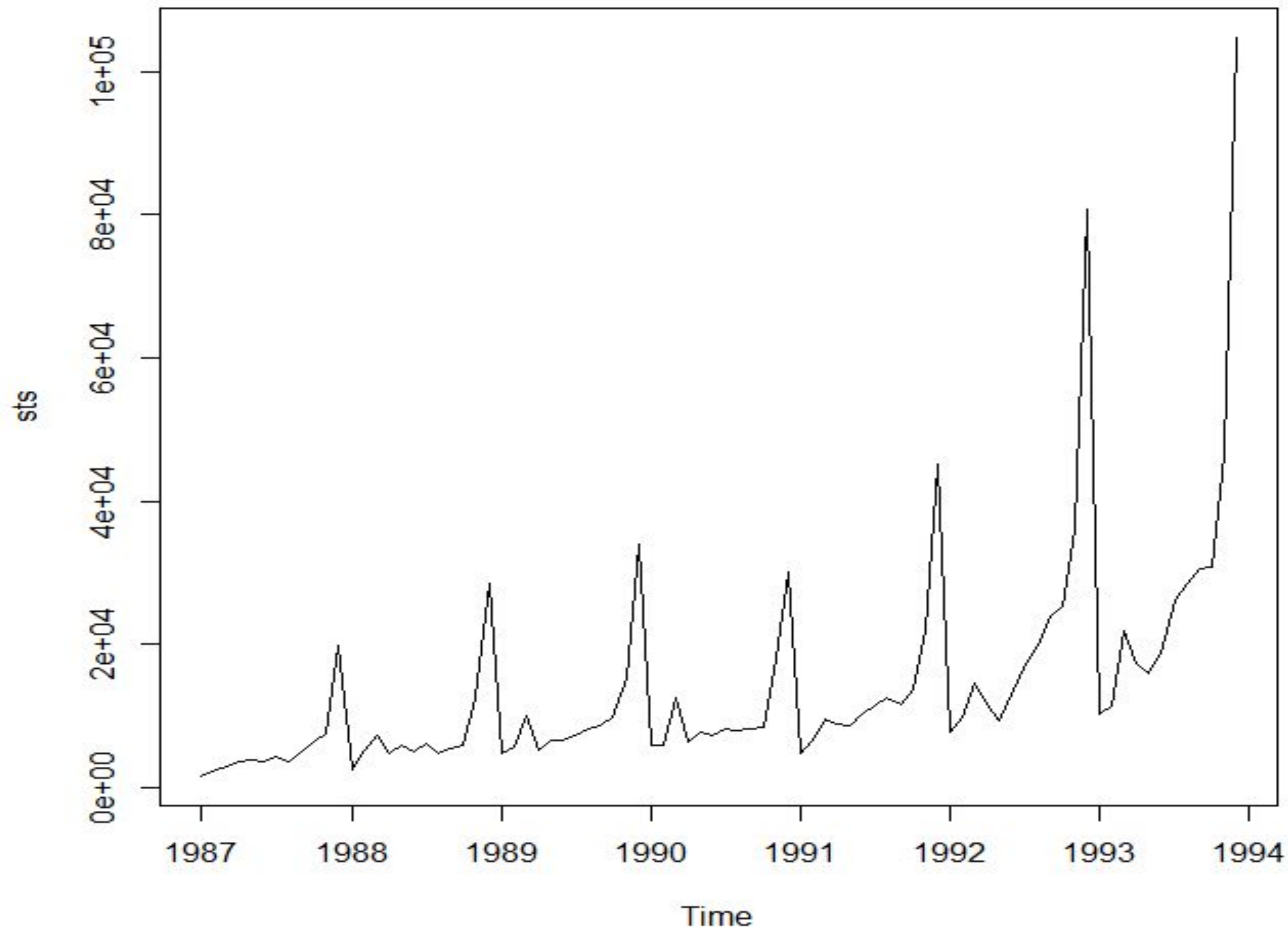


2. Multiplicative Model:-

The multiplicative model accounts for the value of Y at time t as a product of the individual components.



Example of multiplicative Model



2.1 TIME SERIES DECOMPOSITION

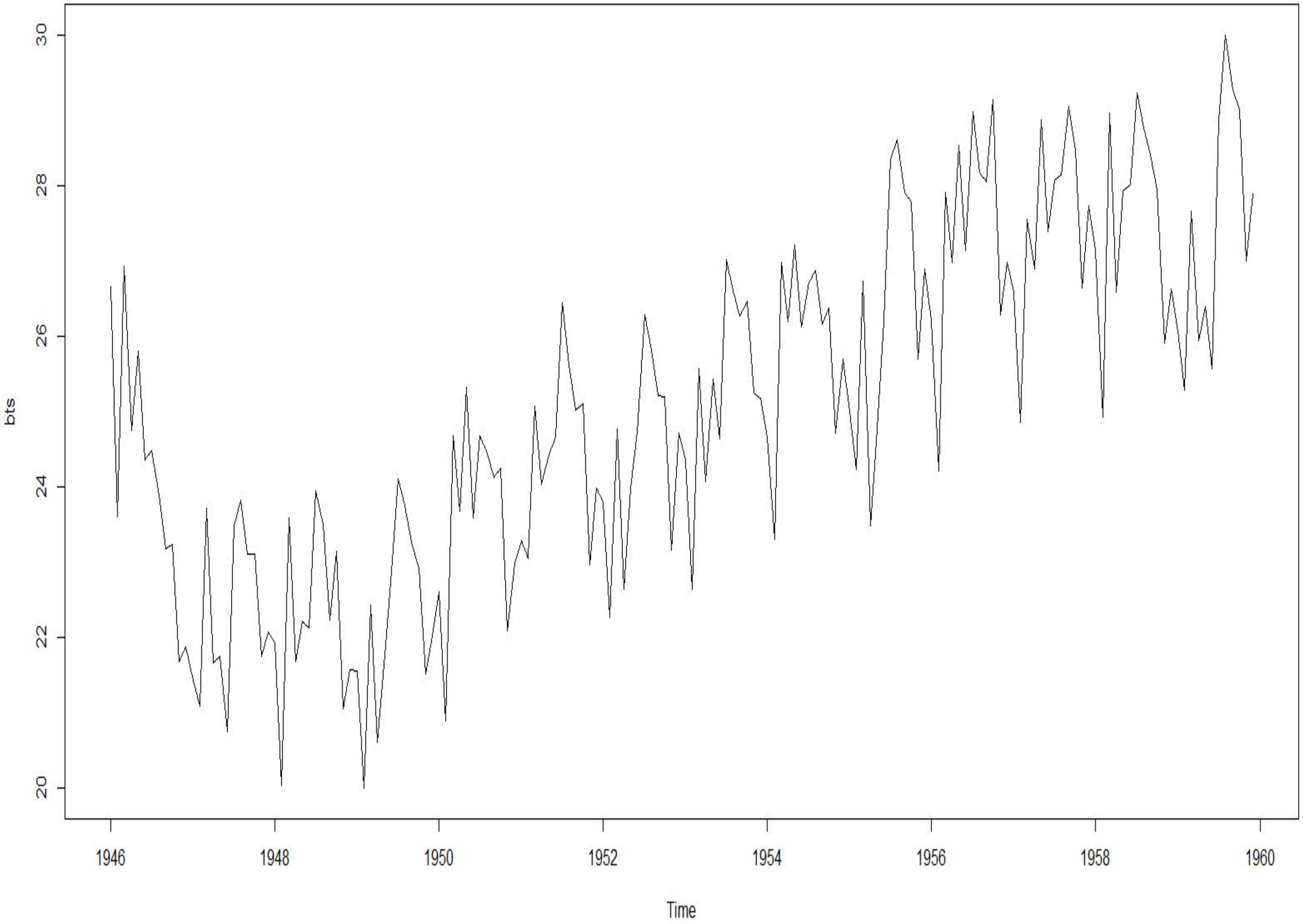
Decomposition of Time Series

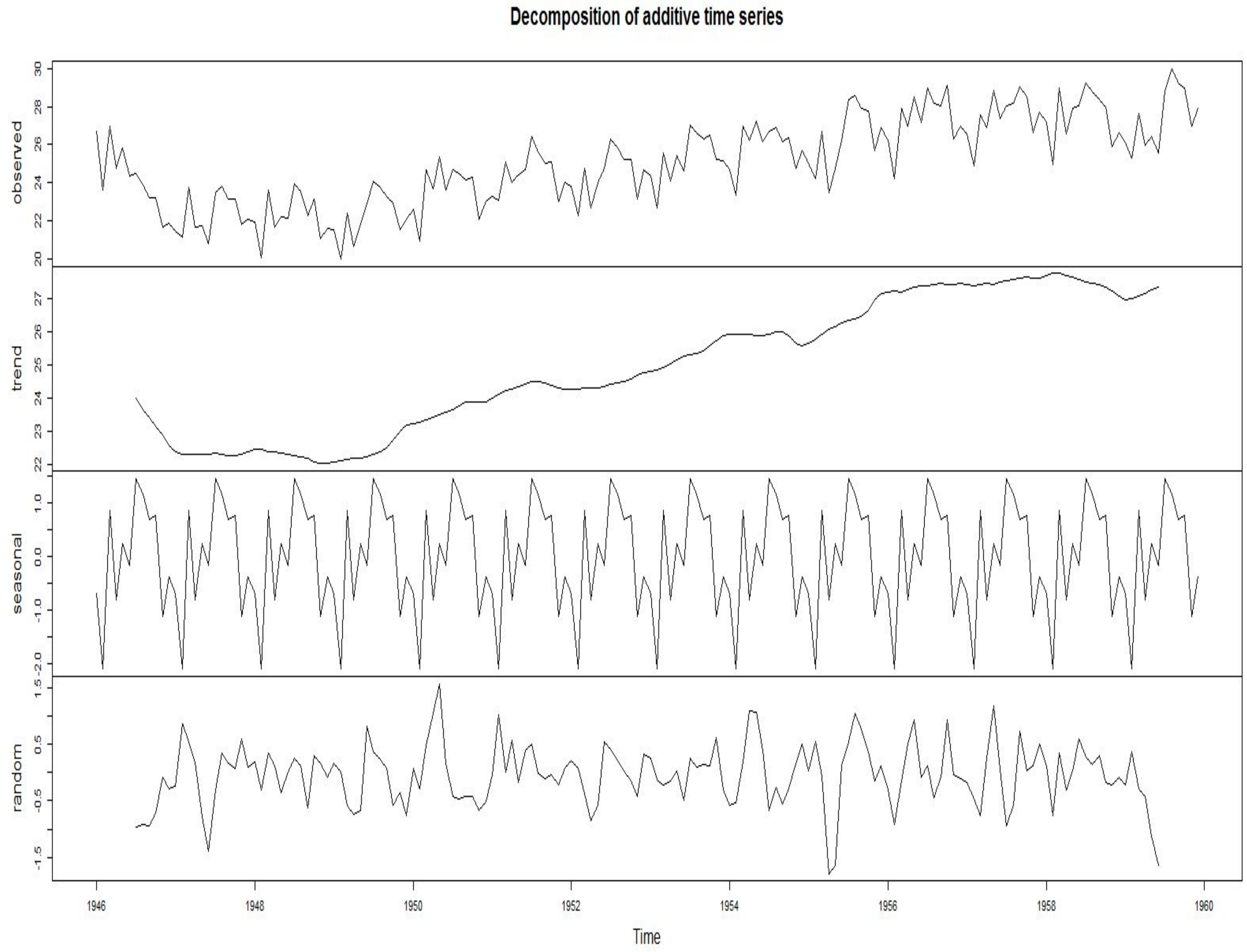
1. A time series decomposition is a mathematical procedure which transform a time series into multiple different time series.

2. The original time series is often decompose into 3 sub-time series:

- Trend
- Seasonal
- Random

Example for Decomposition




Decomposition of additive time series
observed
trend
seasonal
random
Time
1946
1948
1950
1952
1954
1956
1958
1960

Step-by-step: Time series decomposition

Step 1: Detect the trend

Step 2: Detrend the time series

Step 3: Average seasonality

Step 4: Random noise left

2.2 DECOMPOSITION OF ADDITIVE TIME SERIES

Step 1: detect the trend

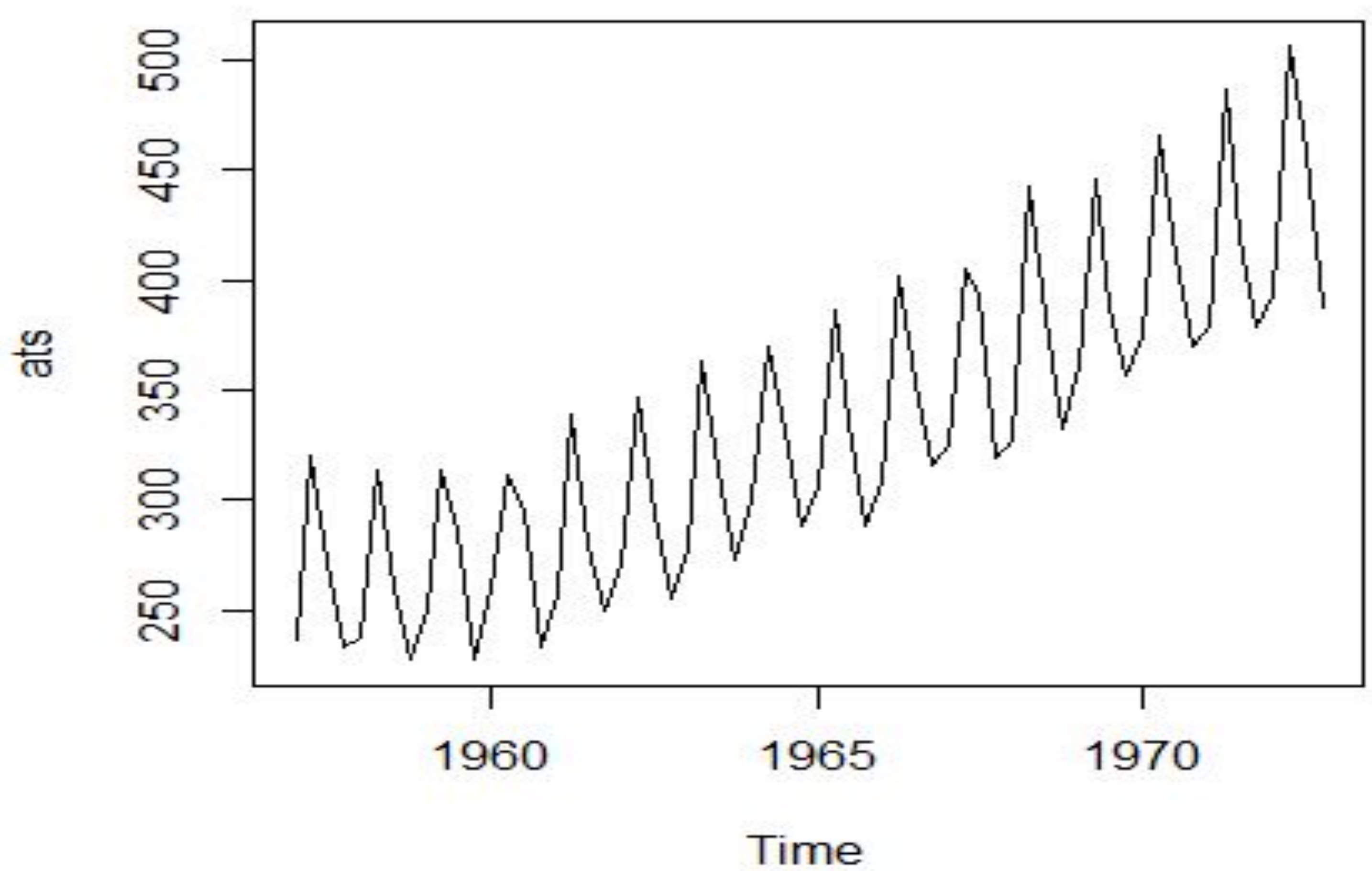
- Trend in a time series is calculated by **Centred Moving Average**

$$T_t = 1/m \sum_{j=-k}^{k} y_{t+j}$$

Where m= 2k+1

Data Set :-

236	320	272	233	237	313	261	227	250	314	286
227	260	311	295	233	257	339	279	250	270	346
294	255	278	363	313	273	300	370	331	288	306
386	335	288	308	402	353	316	325	405	393	319
327	442	383	332	361	446	387	357	374	466	410
370	379	487	419	378	393	506	458	387		


ats
250
300
350
400
450
500
1960
1965
1970
Time

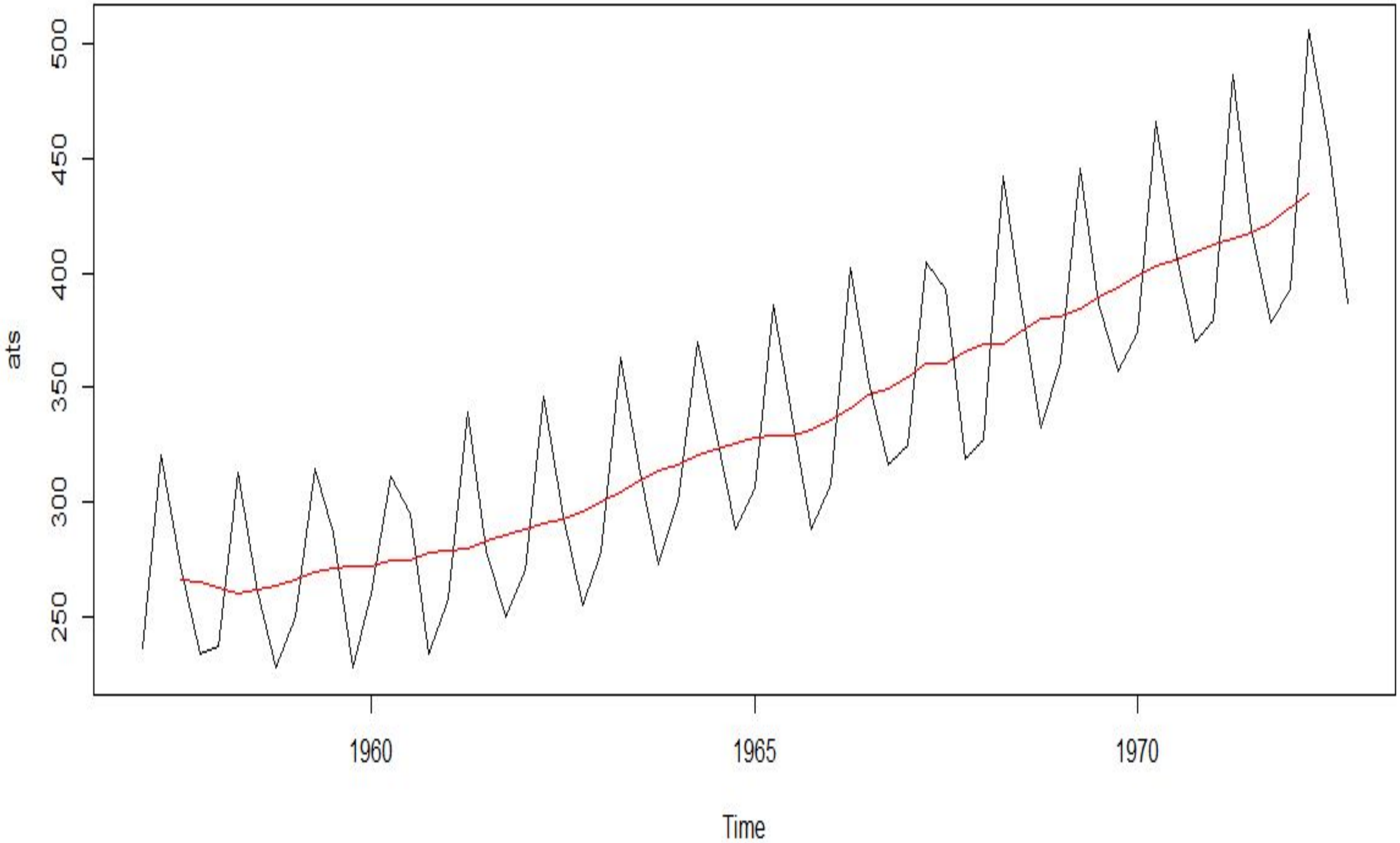
Original Data Set

236	320	272	233	237	313	261	227	250	314	286	227	260
311	295	233	257	339	279	250	270	346	294	255	278	363
313	273	300	370	331	288	306	386	335	288	308	402	353
316	325	405	393	319	327	442	383	332	361	446	387	357
374	466	410	370	379	487	419	378	393	506	458	387	

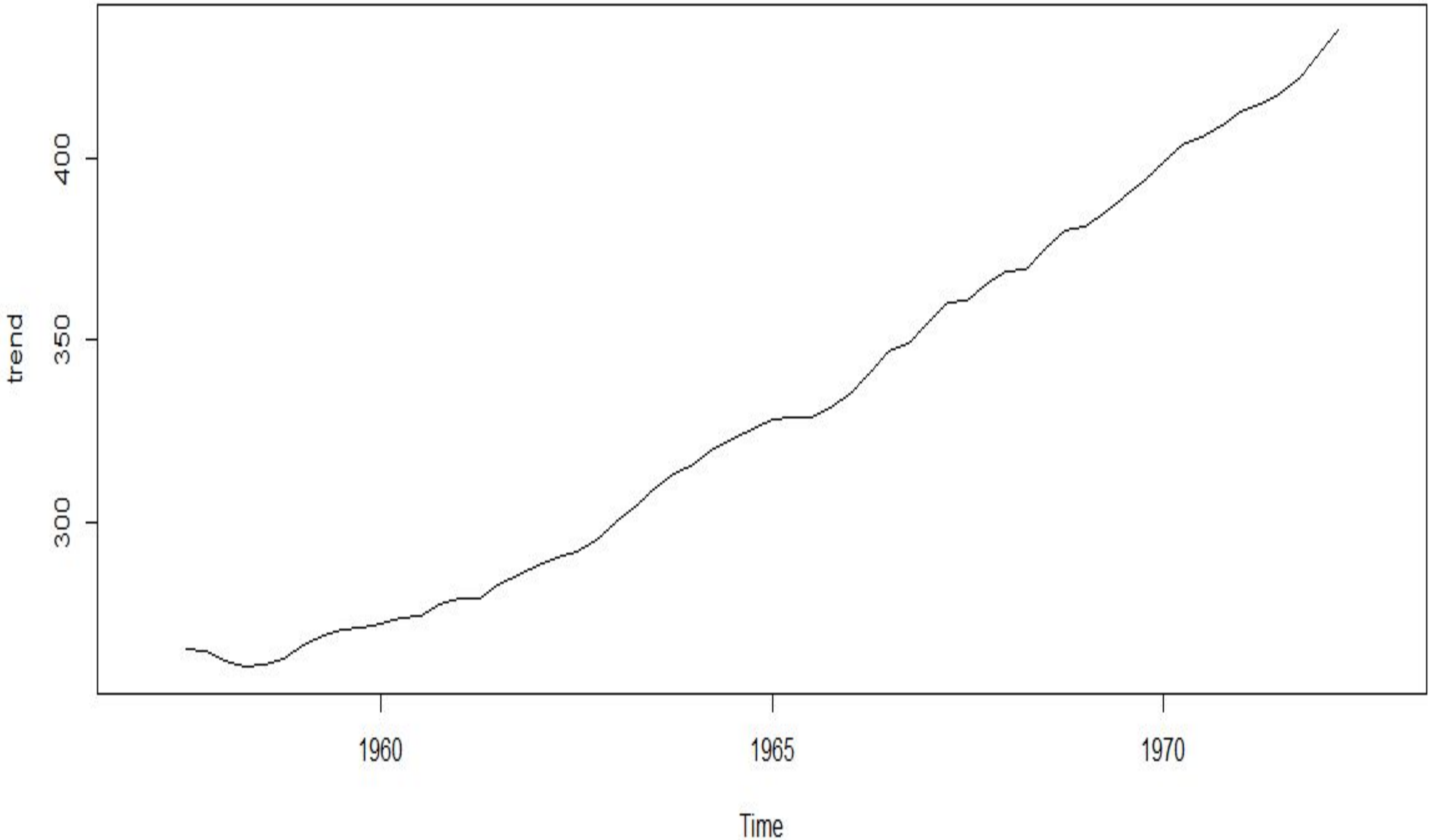
Data Set after Centered Moving Average

NA	NA	265.375	264.625	262.375	260.250	261.125	262.875
266.125	269.250	270.500	271.375	272.125	274.000	274.375	277.500
279.000	279.125	282.875	285.375	288.125	290.625	292.250	295.375
299.875	304.500	309.500	313.125	316.250	320.375	323.000	325.750
328.250	328.750	329.000	331.250	335.500	341.250	346.875	349.375
354.750	360.125	360.750	365.625	369.000	369.375	375.250	380.000
381.000	384.625	389.375	393.500	398.875	403.375	405.625	408.875
412.625	414.750	417.500	421.625	428.875	434.875	NA	NA

Trend



Trend component of time series



Step 2: Detrend the time series

Additive:

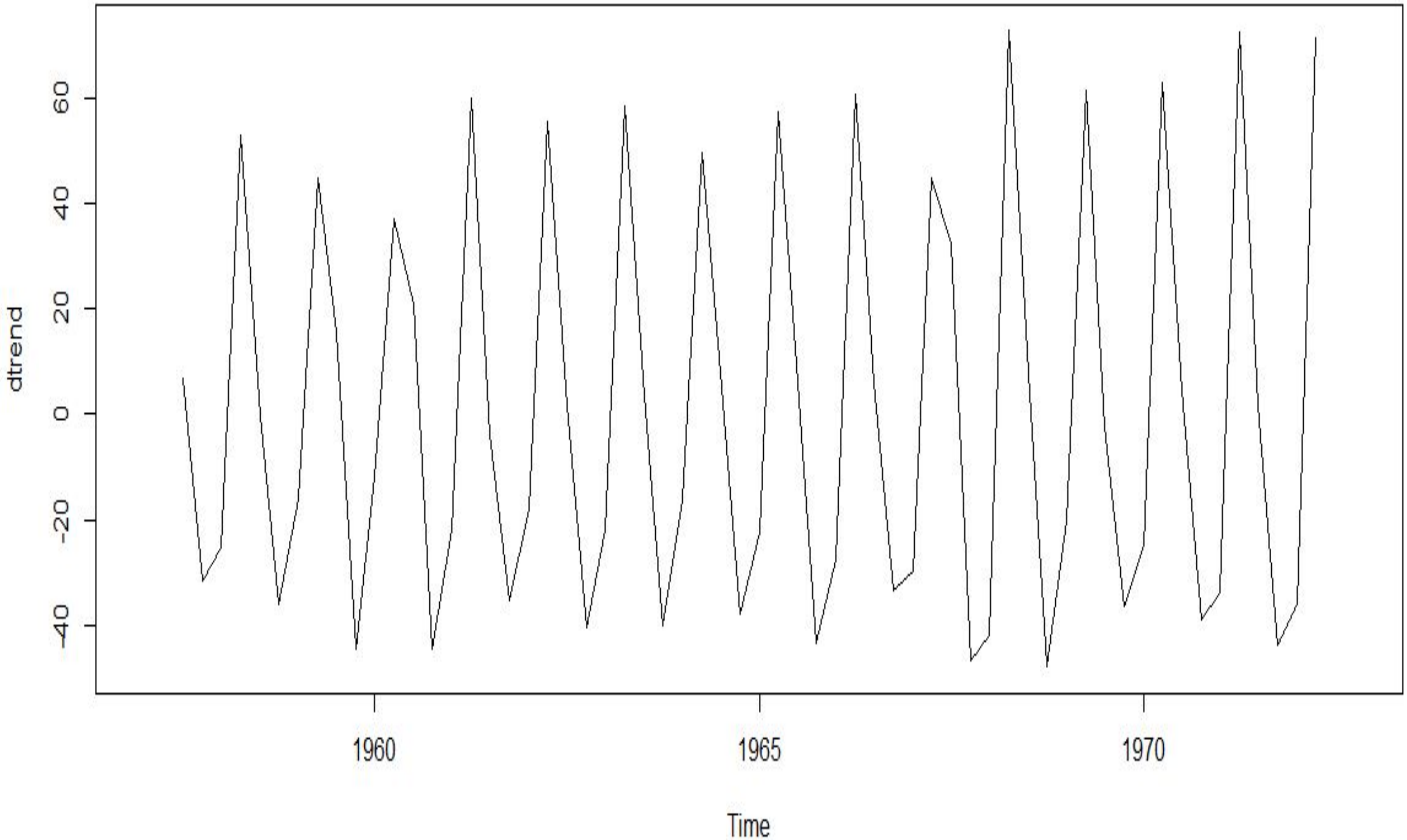
Time series = Seasonal + Trend + Random

Therefore Detrend = Time series - Trend

Original Data set – trend data set

NA	NA	6.625	-31.625	-25.375	52.750	-0.125	-35.875
-16.125	44.750	15.500	-44.375	-12.125	37.000	20.625	-44.500
-22.000	59.875	-3.875	-35.375	-18.125	55.375	1.750	-40.375
-21.875	58.500	3.500	-40.125	-16.250	49.625	8.000	-37.750
-22.250	57.250	6.000	-43.250	-27.500	60.750	6.125	-33.375
-29.750	44.875	32.250	-46.625	-42.000	72.625	7.750	-48.000
-20.000	61.375	-2.375	-36.500	-24.875	62.625	4.375	-38.875
-33.625	72.250	1.500	-43.625	-35.875	71.125	NA	NA

Detrend



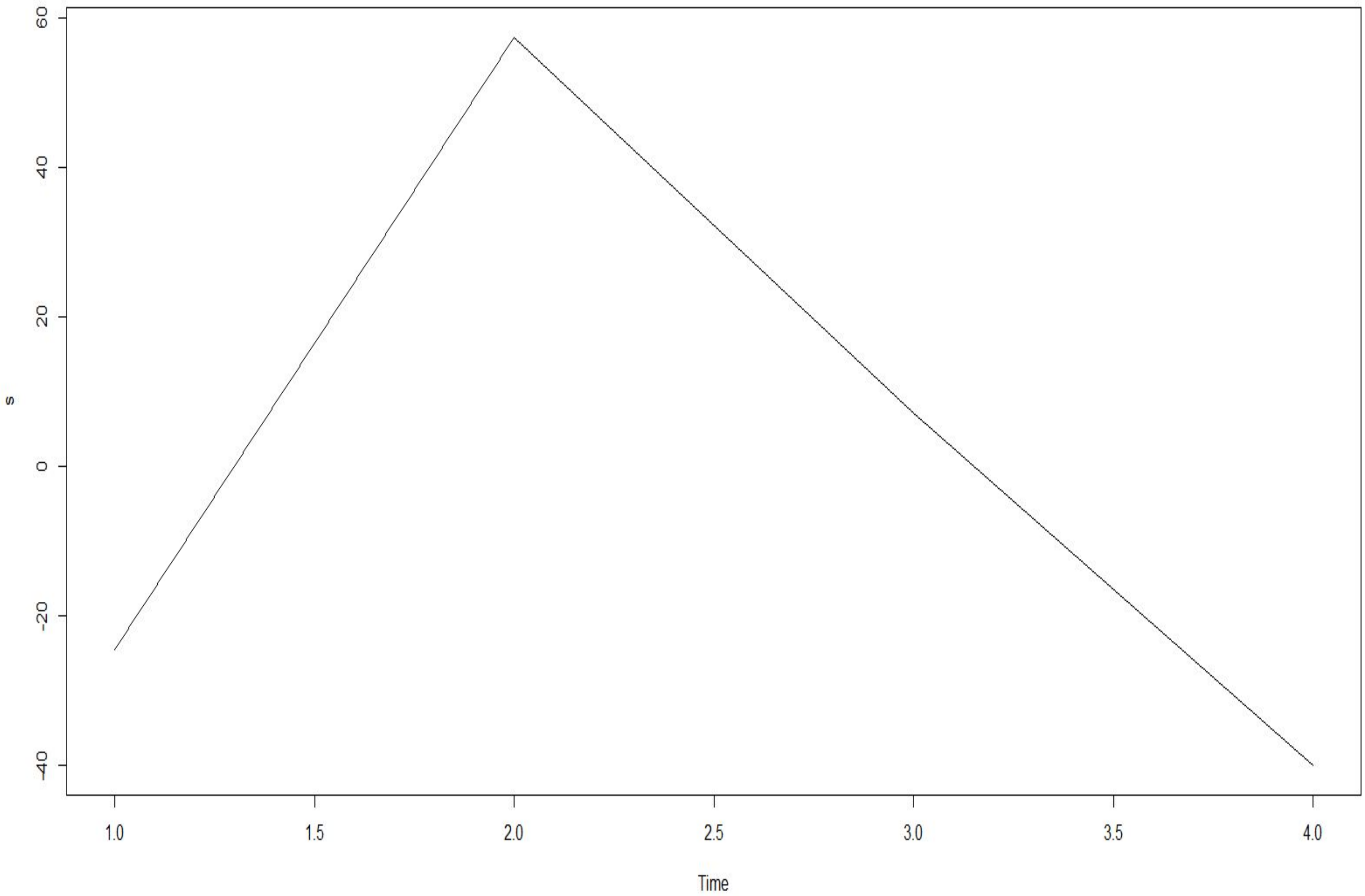
Step 3: Average seasonality

	[,1]	[,2]	[,3]	[,4]
[1,]	NA	NA	6.625	-31.625
[2,]	-25.375	52.750	-0.125	-35.875
[3,]	-16.125	44.750	15.500	-44.375
[4,]	-12.125	37.000	20.625	-44.500
[5,]	-22.000	59.875	-3.875	-35.375
[6,]	-18.125	55.375	1.750	-40.375
[7,]	-21.875	58.500	3.500	-40.125
[8,]	-16.250	49.625	8.000	-37.750
[9,]	-22.250	57.250	6.000	-43.250
[10,]	-27.500	60.750	6.125	-33.375
[11,]	-29.750	44.875	32.250	-46.625
[12,]	-42.000	72.625	7.750	-48.000
[13,]	-20.000	61.375	-2.375	-36.500
[14,]	-24.875	62.625	4.375	-38.875
[15,]	-33.625	72.250	1.500	-43.625
[16,]	-35.875	71.125	NA	NA

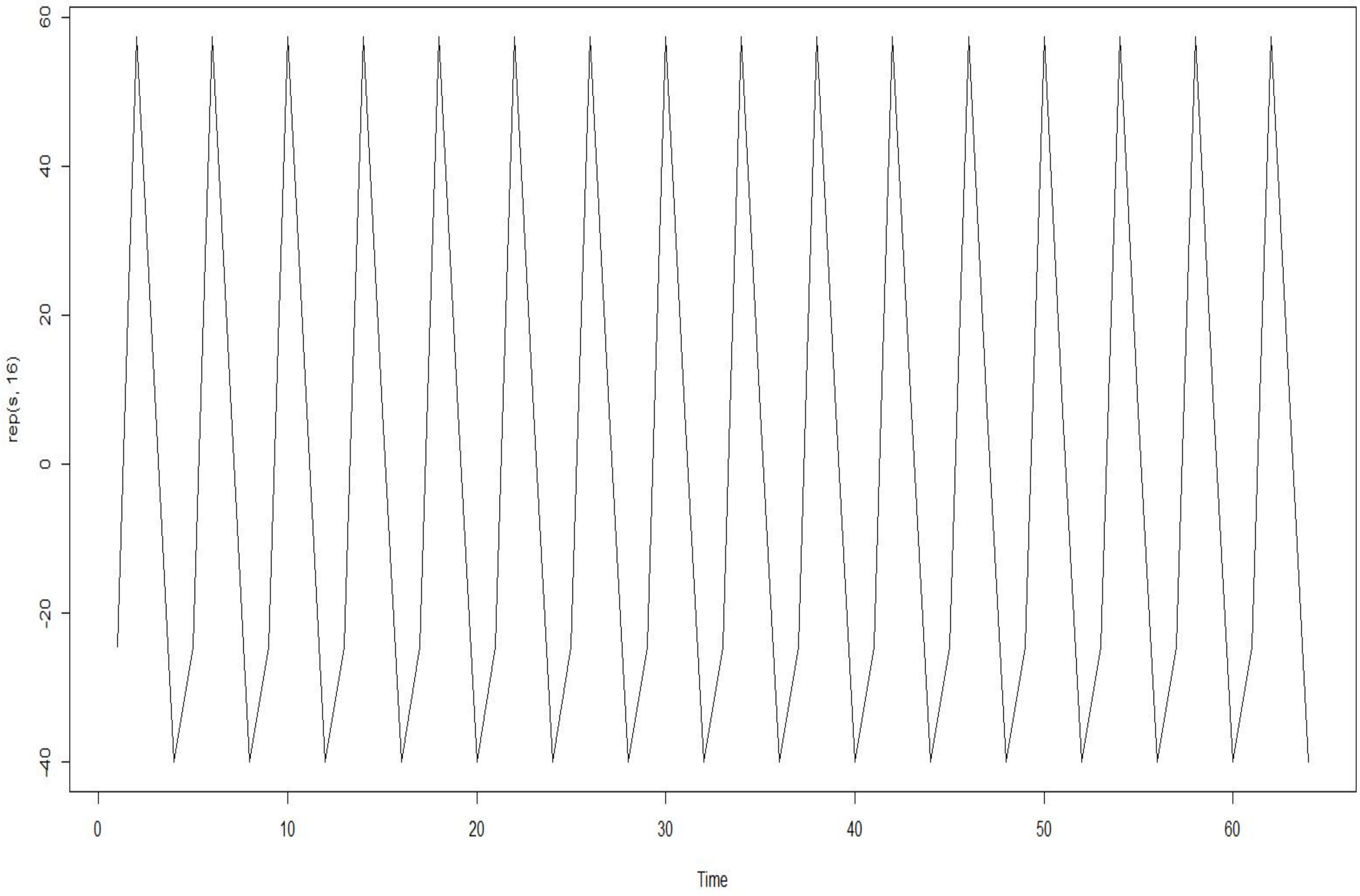
Mean of column

-24.51667 57.38333 7.17500 -40.01667

Seasonal component of time series



Seasonal component of time series



Step 4: Random noise left

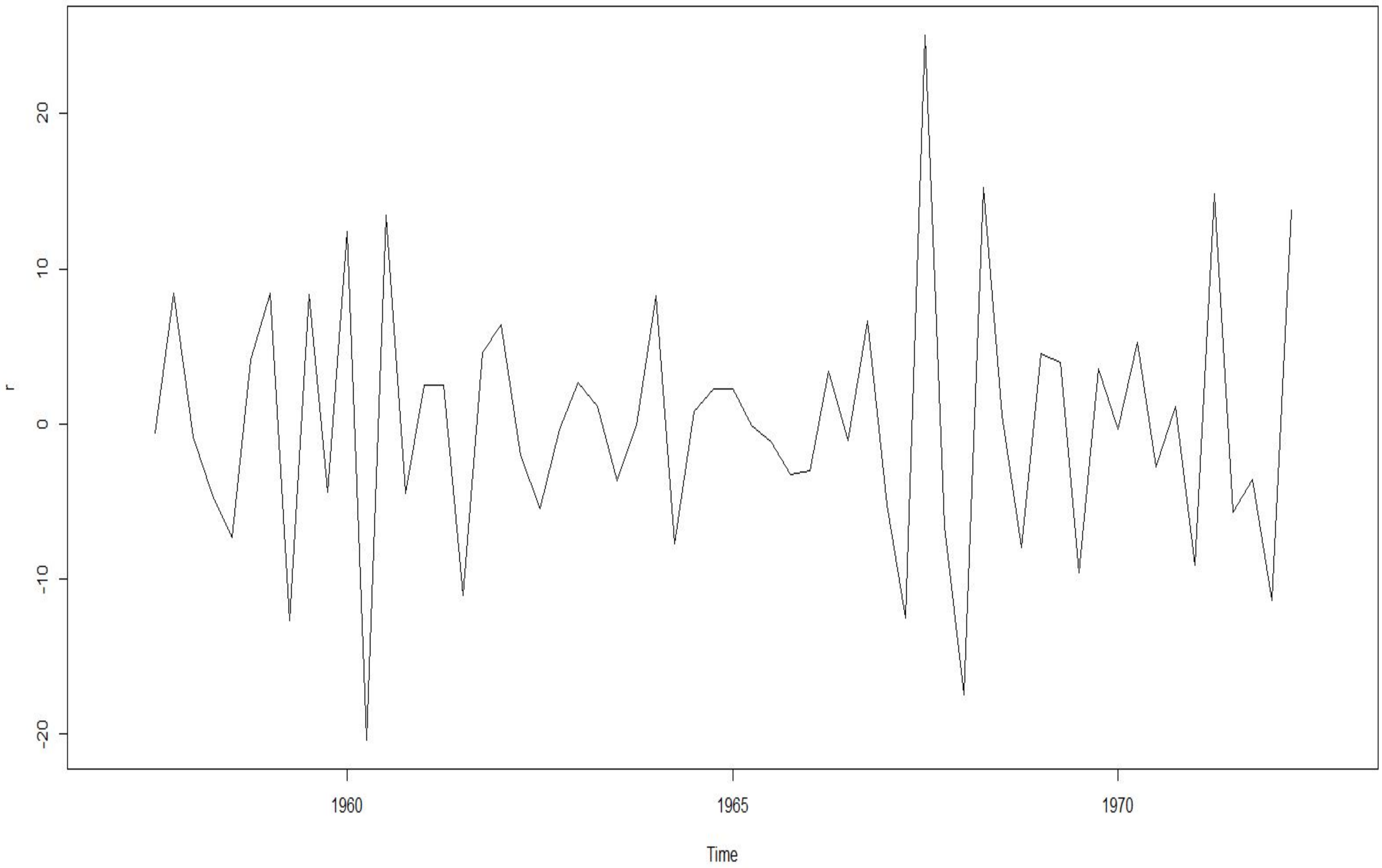
Time series = Seasonal + Trend + Random

Random = Time series - Seasonal - Trend

Random = Time series - Seasonal - Trend

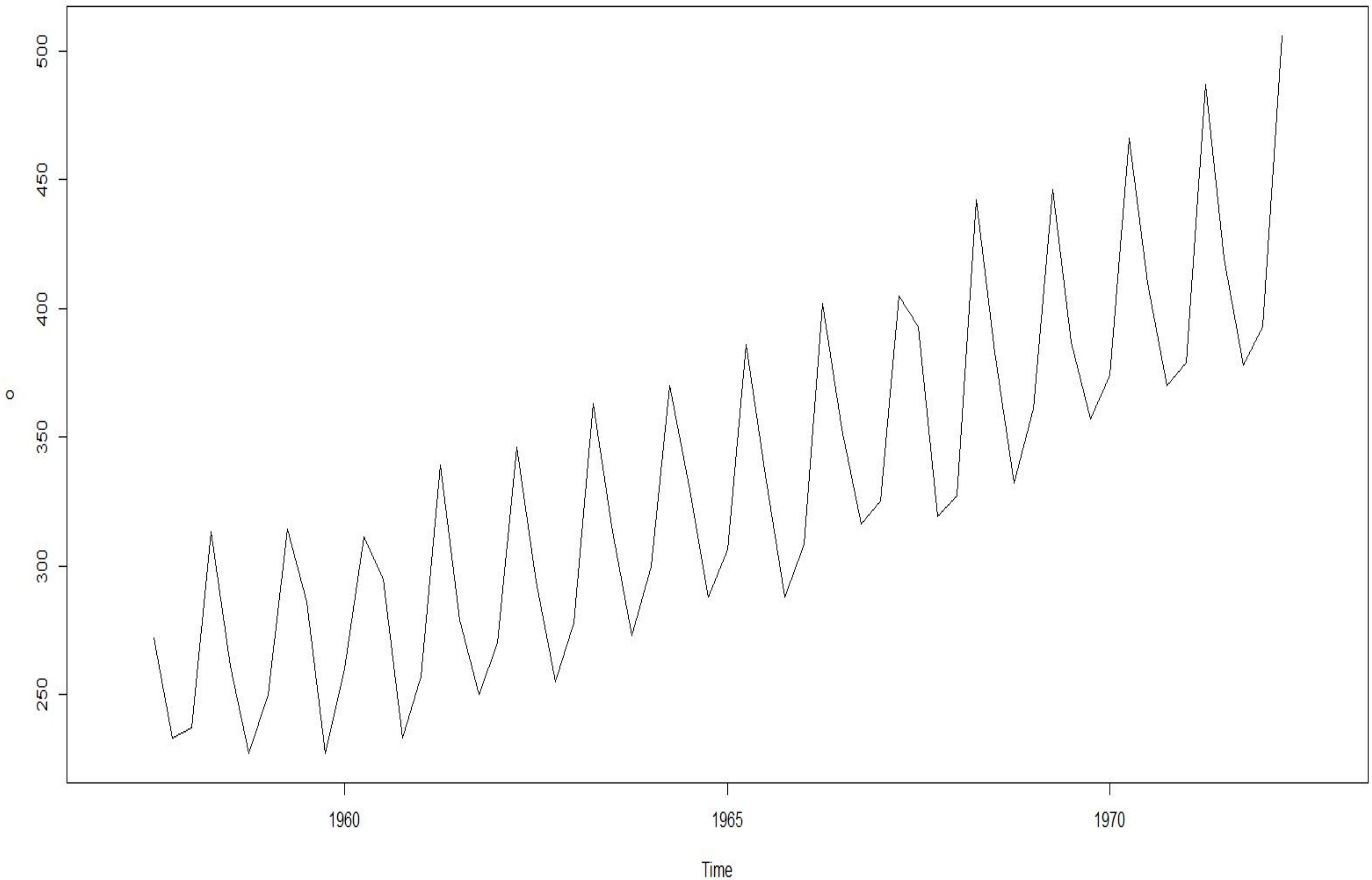
NA	NA	-0.5500000	8.3916667
-0.8583333	-4.6333333	-7.3000000	4.1416667
8.3916667	-12.6333333	8.3250000	-4.3583333
12.3916667	-20.3833333	13.4500000	-4.4833333
2.5166667	2.4916667	-11.0500000	4.6416667
6.3916667	-2.0083333	-5.4250000	-0.3583333
2.6416667	1.1166667	-3.6750000	-0.1083333
8.2666667	-7.7583333	0.8250000	2.2666667
2.2666667	-0.1333333	-1.1750000	-3.2333333
-2.9833333	3.3666667	-1.0500000	6.6416667
-5.2333333	-12.5083333	25.0750000	-6.6083333
-17.4833333	15.2416667	0.5750000	-7.9833333
4.5166667	3.9916667	-9.5500000	3.5166667
-0.3583333	5.2416667	-2.8000000	1.1416667
-9.1083333	14.8666667	-5.6750000	-3.6083333
-11.3583333	13.7416667	NA	NA

Random component of time series



Time series = Seasonal + Trend + Random

NA NA 272 233 237 313 261 227 250 314 286 227 260 311
295 233 257 339 279 250 270 346 294 255 278 363 313 273
300 370 331 288 306 386 335 288 308 402 353 316 325 405
393 319 327 442 383 332 361 446 387 357 374 466 410 370
379 487 419 378 393 506 NA NA



500
450
400
350
300
250
o
1960
1965
1970
Time

2.3 DECOMPOSITION OF MULTIPLICATIVE TIME SERIES

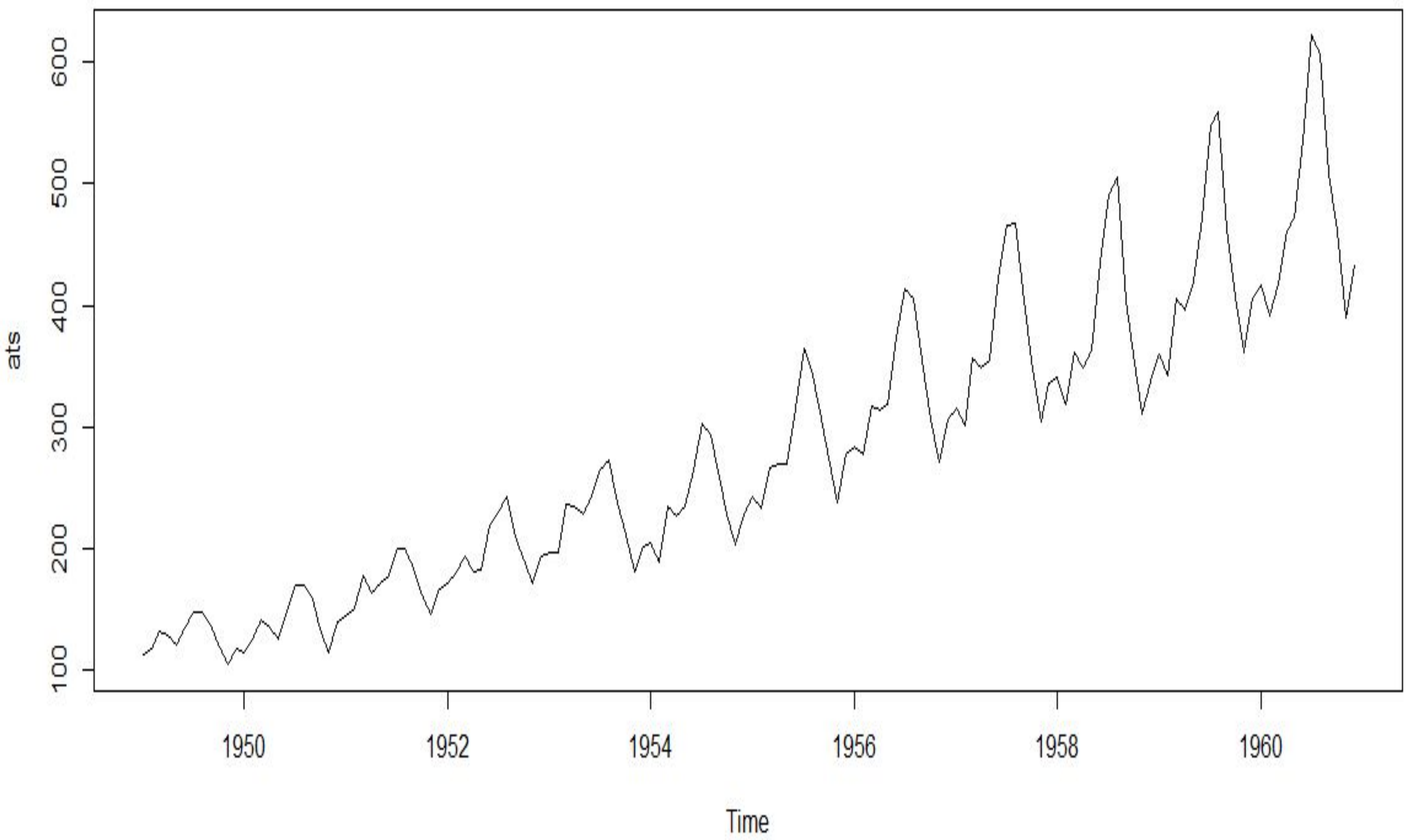
Step 1: detect the trend

- Trend in a time series is calculated by **Centred Moving Average** same as in the additive model

$$T_t = 1/m \sum_{j=-k}^{k} y_{t+j}$$

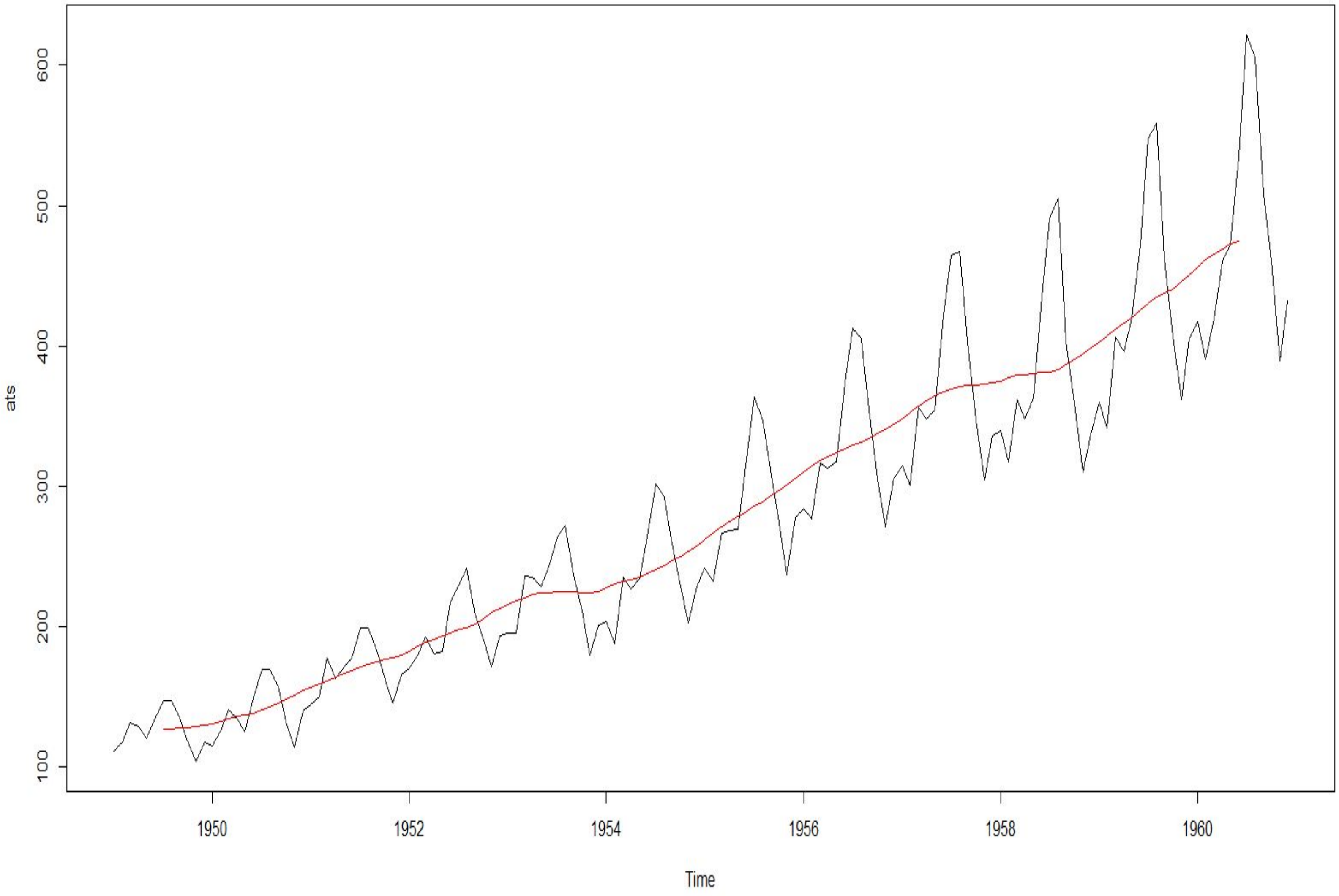
Data Set

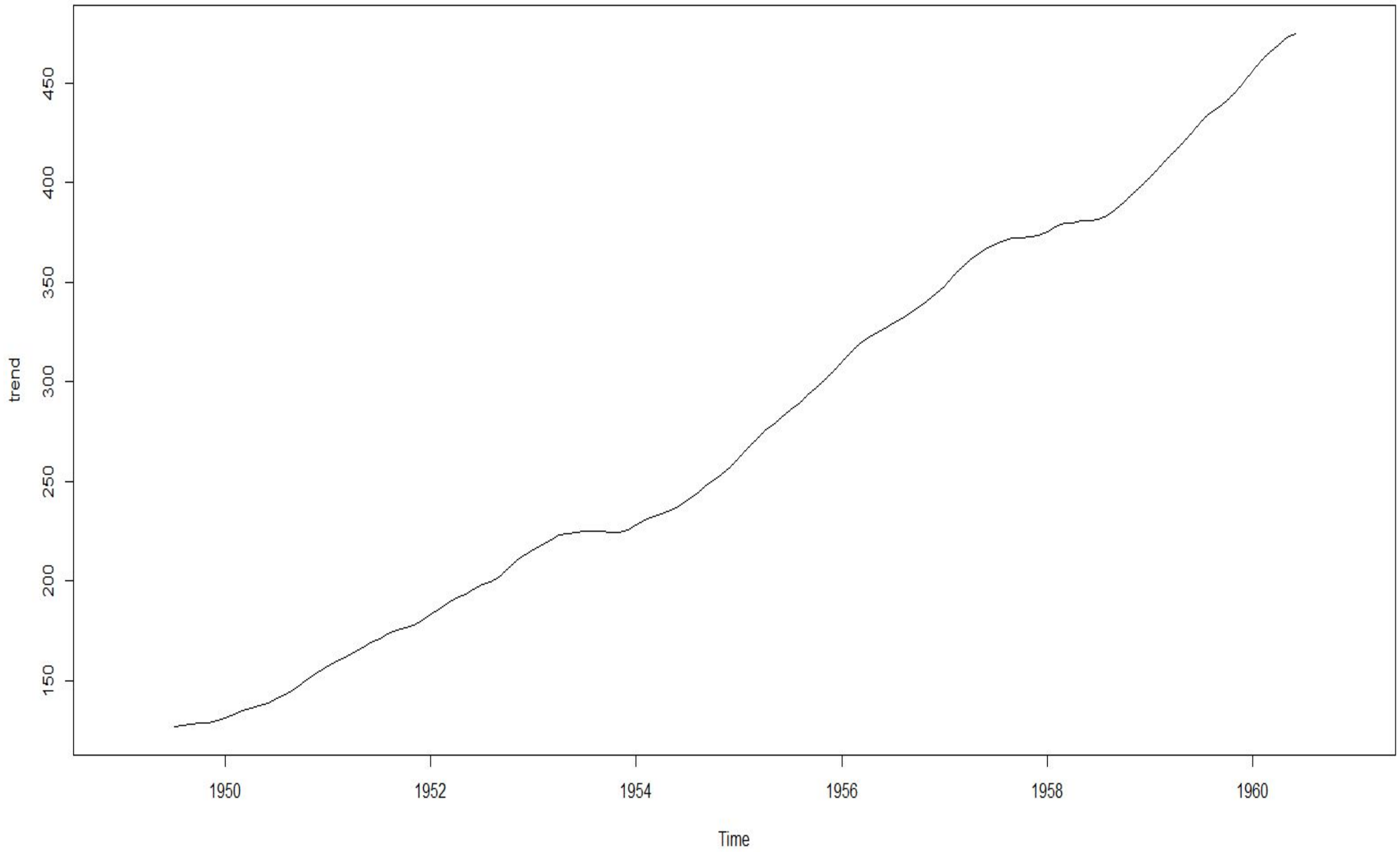
112	118	132	129	121	135	148	148	136	119	104	118
115	126	141	135	125	149	170	170	158	133	114	140
145	150	178	163	172	178	199	199	184	162	146	166
171	180	193	181	183	218	230	242	209	191	172	194
196	196	236	235	229	243	264	272	237	211	180	201
204	188	235	227	234	264	302	293	259	229	203	229
242	233	267	269	270	315	364	347	312	274	237	278
284	277	317	313	318	374	413	405	355	306	271	306
315	301	356	348	355	422	465	467	404	347	305	336
340	318	362	348	363	435	491	505	404	359	310	337
360	342	406	396	420	472	548	559	463	407	362	405
417	391	419	461	472	535	622	606	508	461	390	432



ats
100
200
300
400
500
600
1950
1952
1954
1956
1958
1960
Time

```
      NA       NA       NA       NA       NA       NA 126.7917 127.2500 127.9583
128.5833 129.0000 129.7500 131.2500 133.0833 134.9167 136.4167
137.4167 138.7500 140.9167 143.1667 145.7083 148.4167 151.5417
154.7083 157.1250 159.5417 161.8333 164.1250 166.6667 169.0833
171.2500 173.5833 175.4583 176.8333 178.0417 180.1667 183.1250
186.2083 189.0417 191.2917 193.5833 195.8333 198.0417 199.7500
202.2083 206.2500 210.4167 213.3750 215.8333 218.5000 220.9167
222.9167 224.0833 224.7083 225.3333 225.3333 224.9583 224.5833
224.4583 225.5417 228.0000 230.4583 232.2500 233.9167 235.6250
237.7500 240.5000 243.9583 247.1667 250.2500 253.5000 257.1250
261.8333 266.6667 271.1250 275.2083 278.5000 281.9583 285.7500
289.3333 293.2500 297.1667 301.0000 305.4583 309.9583 314.4167
318.6250 321.7500 324.5000 327.0833 329.5417 331.8333 334.4583
337.5417 340.5417 344.0833 348.2500 353.0000 357.6250 361.3750
364.5000 367.1667 369.4583 371.2083 372.1667 372.4167 372.7500
373.6250 375.2500 377.9167 379.5000 380.0000 380.7083 380.9583
381.8333 383.6667 386.5000 390.3333 394.7083 398.6250 402.5417
407.1667 411.8750 416.3333 420.5000 425.5000 430.7083 435.1250
437.7083 440.9583 445.8333 450.6250 456.3333 461.3750 465.2083
469.3333 472.7500 475.0417       NA       NA       NA       NA       NA       NA
```


600
500
400
300
200
100
ats
1950
1952
1954
1956
1958
1960
Time



trend
150
200
250
300
350
400
450
1950
1952
1954
1956
1958
1960
Time

Step 2: Detrend the time series

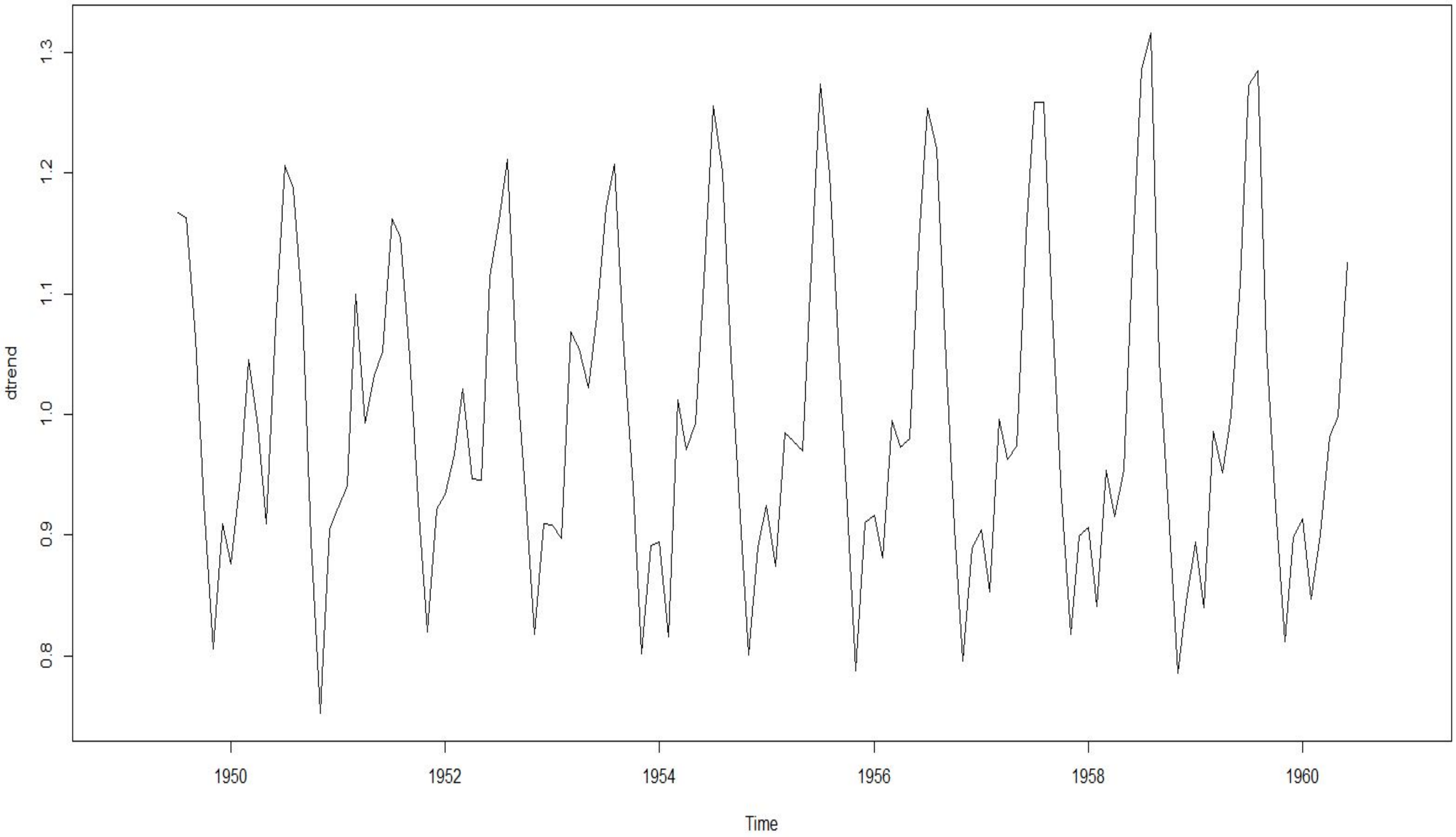
Additive:

Time series = Seasonal * Trend * Random

Therefore Detrend = Time series / Trend

Original Data set / trend data set

NA NA NA NA NA NA 1.1672691 1.1630648 1.0628460
0.9254699 0.8062016 0.9094412 0.8761905 0.9467752 1.0450896 0.9896151
0.9096422 1.0738739 1.2063868 1.1874272 1.0843580 0.8961258 0.7522684
0.9049286 0.9228321 0.9401933 1.0998970 0.9931455 1.0320000 1.0527353
1.1620438 1.1464234 1.0486820 0.9161169 0.8200328 0.9213691 0.9337884
0.9666592 1.0209389 0.9461991 0.9453293 1.1131915 1.1613718 1.2115144
1.0335875 0.9260606 0.8174257 0.9091974 0.9081081 0.8970252 1.0682761
1.0542056 1.0219412 1.0814018 1.1715976 1.2071006 1.0535284 0.9395176
0.8019306 0.8911879 0.8947368 0.8157657 1.0118407 0.9704311 0.9931034
1.1104101 1.2557173 1.2010248 1.0478759 0.9150849 0.8007890 0.8906174
0.9242521 0.8737500 0.9847856 0.9774413 0.9694794 1.1171863 1.2738408
1.1993088 1.0639386 0.9220415 0.7873754 0.9101078 0.9162522 0.8809966
0.9949000 0.9728050 0.9799692 1.1434395 1.2532558 1.2204922 1.0614177
0.9065547 0.7957910 0.8893194 0.9045226 0.8526912 0.9954561 0.9629886
0.9739369 1.1493418 1.2585993 1.2580537 1.0855352 0.9317521 0.8182428
0.8992974 0.9060626 0.8414553 0.9538867 0.9157895 0.9534858 1.1418572
1.2859014 1.3162467 1.0452781 0.9197267 0.7853901 0.8454061 0.8943174
0.8399509 0.9857360 0.9511609 0.9988109 1.1092832 1.2723227 1.2846883
1.0577820 0.9229897 0.8119626 0.8987517 0.9138057 0.8474668 0.9006717
0.9822443 0.9984135 1.1262170 NA NA NA NA NA NA


1.3
1.2
1.1
1.0
0.9
0.8
dtrend
1950
1952
1954
1956
1958
1960
Time

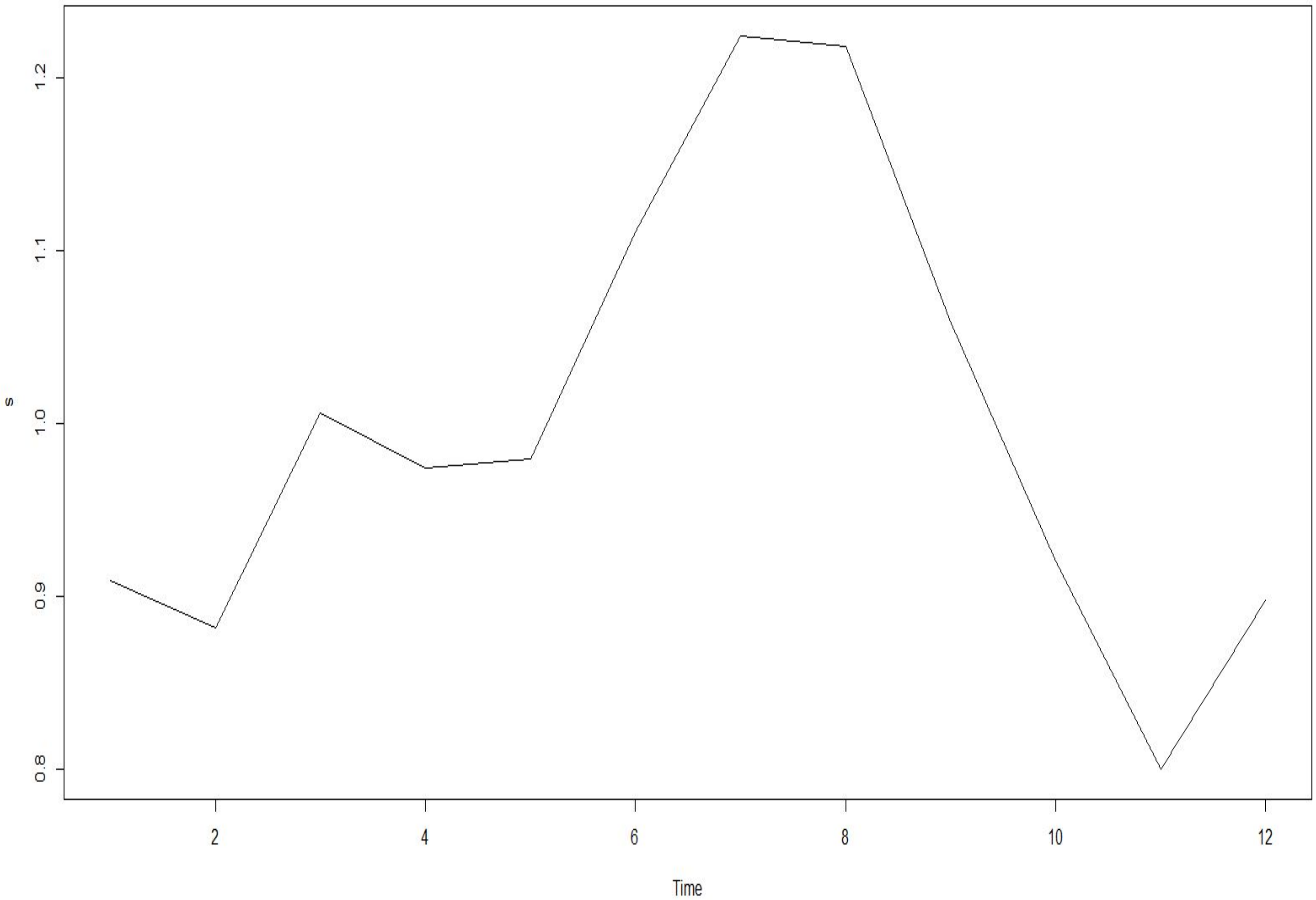
Step 3: Average seasonality

NA	NA	NA	NA	NA	NA	1.167269	1.163065	1.062846	0.9254699	0.8062016	0.9094412
0.8761905	0.9467752	1.0450896	0.9896151	0.9096422	1.073874	1.206387	1.187427	1.084358	0.8961258	0.7522684	0.9049286
0.9228321	0.9401933	1.0998970	0.9931455	1.0320000	1.052735	1.162044	1.146423	1.048682	0.9161169	0.8200328	0.9213691
0.9337884	0.9666592	1.0209389	0.9461991	0.9453293	1.113191	1.161372	1.211514	1.033587	0.9260606	0.8174257	0.9091974
0.9081081	0.8970252	1.0682761	1.0542056	1.0219412	1.081402	1.171598	1.207101	1.053528	0.9395176	0.8019306	0.8911879
0.8947368	0.8157657	1.0118407	0.9704311	0.9931034	1.110410	1.255717	1.201025	1.047876	0.9150849	0.8007890	0.8906174
0.9242521	0.8737500	0.9847856	0.9774413	0.9694794	1.117186	1.273841	1.199309	1.063939	0.9220415	0.7873754	0.9101078
0.9162522	0.8809966	0.9949000	0.9728050	0.9799692	1.143439	1.253256	1.220492	1.061418	0.9065547	0.7957910	0.8893194
0.9045226	0.8526912	0.9954561	0.9629886	0.9739369	1.149342	1.258599	1.258054	1.085535	0.9317521	0.8182428	0.8992974
0.9060626	0.8414553	0.9538867	0.9157895	0.9534858	1.141857	1.285901	1.316247	1.045278	0.9197267	0.7853901	0.8454061
0.8943174	0.8399509	0.9857360	0.9511609	0.9988109	1.109283	1.272323	1.284688	1.057782	0.9229897	0.8119626	0.8987517
0.9138057	0.8474668	0.9006717	0.9822443	0.9984135	1.126217	NA	NA	NA	NA	NA	NA

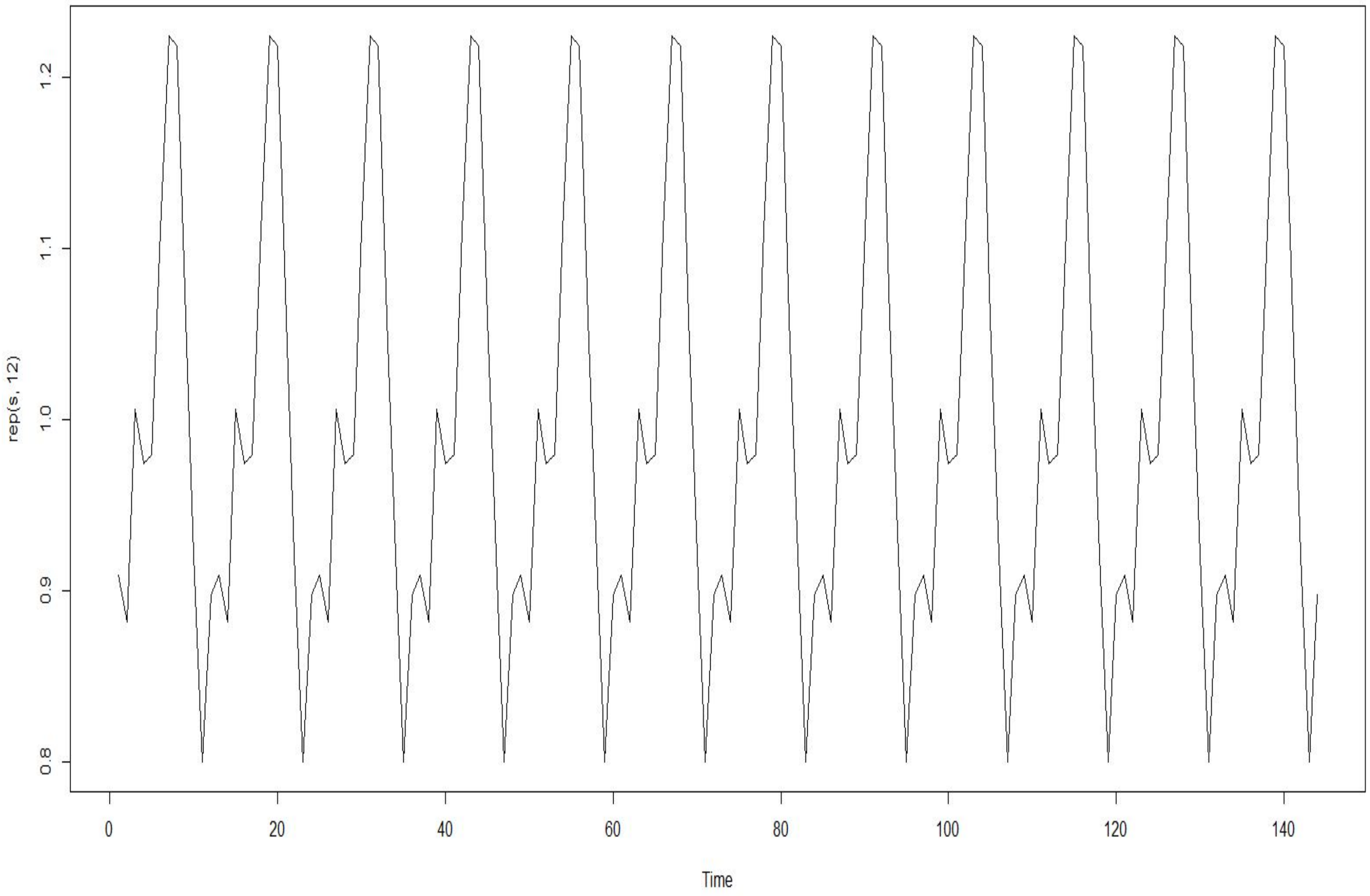
0.9086244 0.8820663 1.0055889 0.9741842 0.9796465 1.1108125 1.2243915 1.2177586 1.0586209 0.9201309 0.7997645 0.8972386

Mean of column

Seasonal component of time series



Seasonal component of time series



Step 4: Random noise left

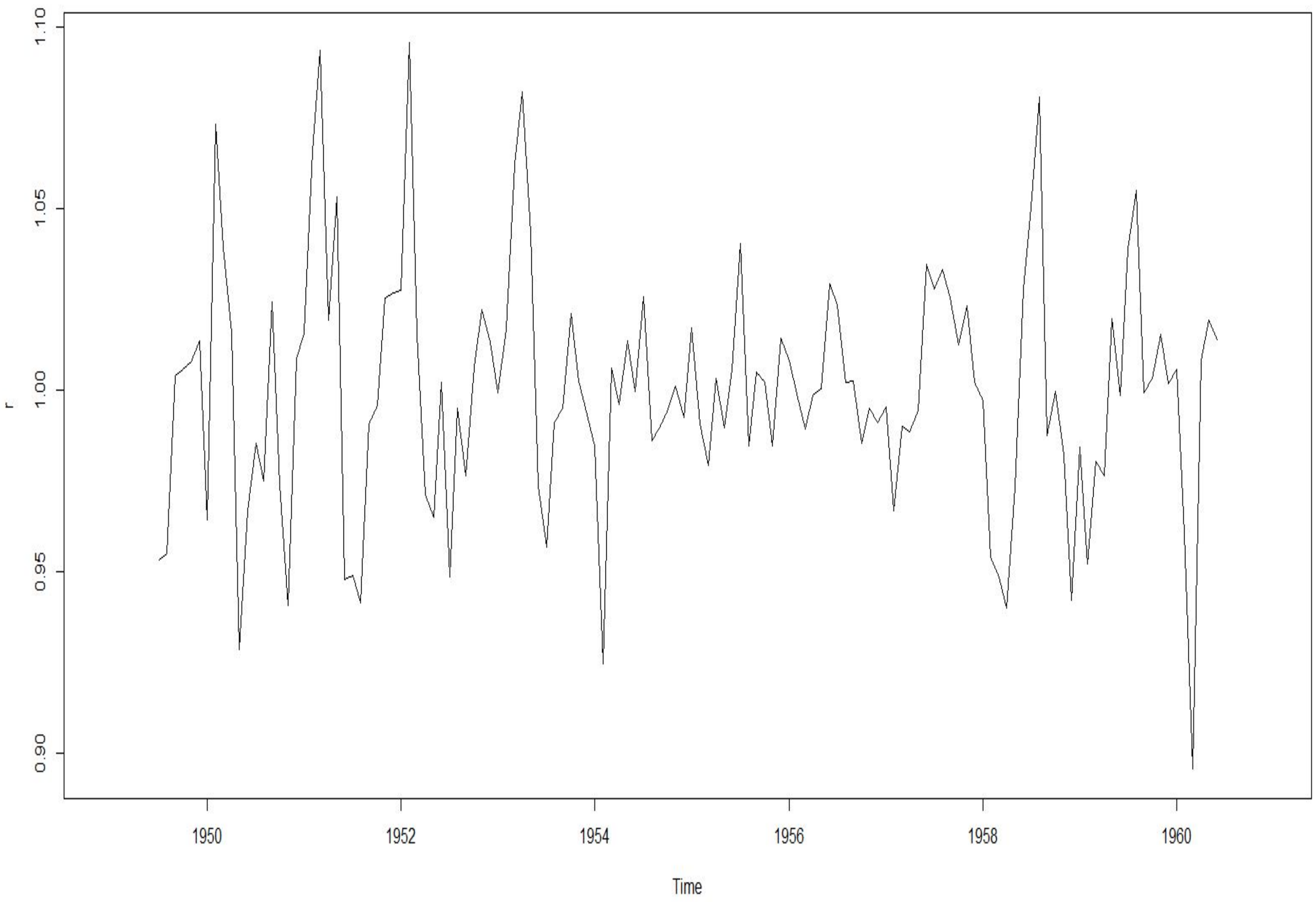
Time series = Seasonal * Trend * Random

Random = Time series / (Seasonal * Trend)

Random = Time series / (Seasonal * Trend)

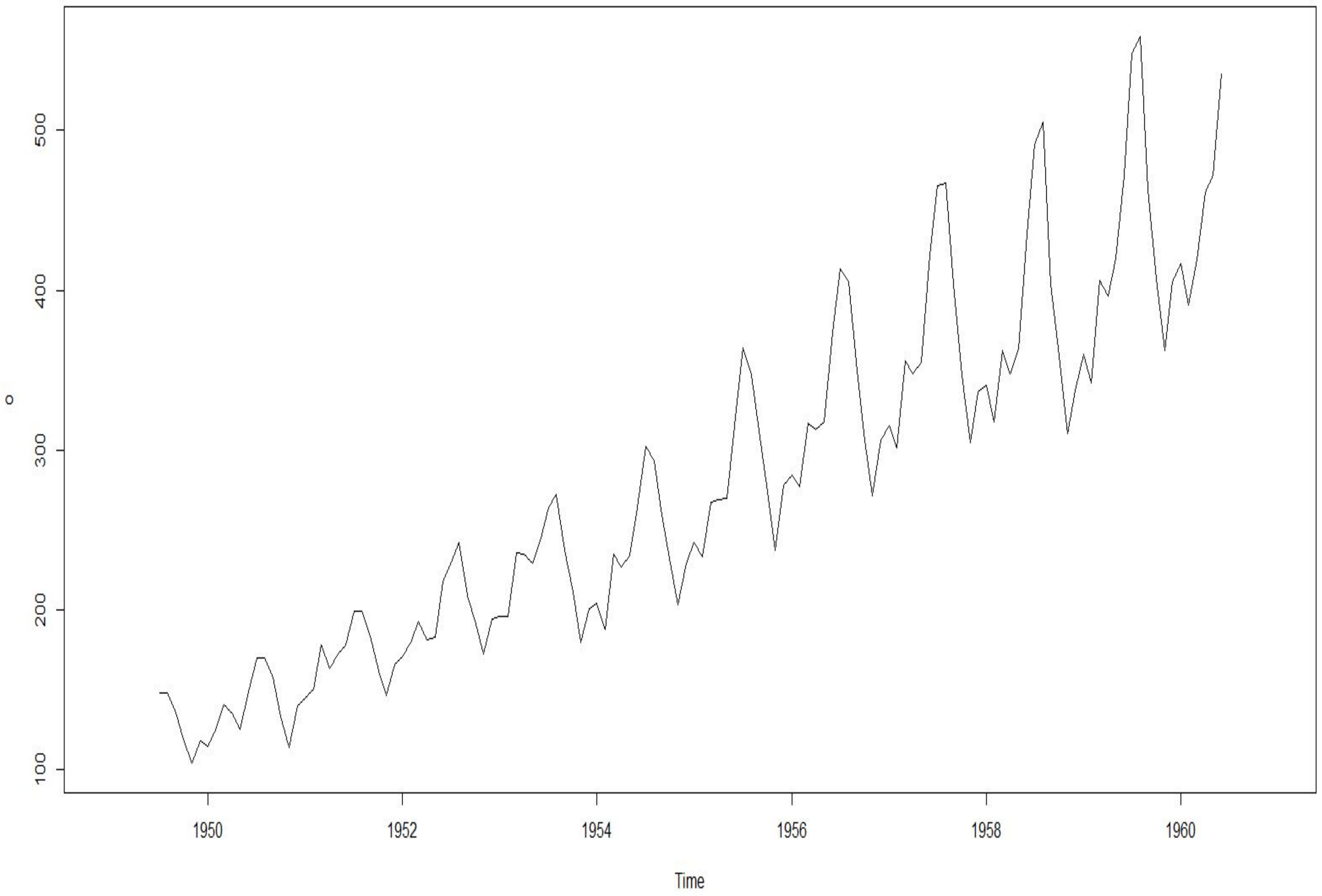
NA	NA	NA	NA	NA	NA	0.9533463	0.9550865	1.0039912
1.0058023	1.0080486	1.0136003	0.9643044	1.0733606	1.0392811	1.0158399	0.9285412	0.9667463
0.9852950	0.9750924	1.0243120	0.9739111	0.9406123	1.0085708	1.0156365	1.0658986	1.0937839
1.0194638	1.0534412	0.9477165	0.9490786	0.9414209	0.9906115	0.9956375	1.0253428	1.0268942
1.0276946	1.0959031	1.0152647	0.9712733	0.9649698	1.0021417	0.9485298	0.9948724	0.9763528
1.0064444	1.0220830	1.0133285	0.9994318	1.0169589	1.0623388	1.0821420	1.0431734	0.9735233
0.9568816	0.9912478	0.9951896	1.0210695	1.0027084	0.9932563	0.9847158	0.9248349	1.0062170
0.9961474	1.0137365	0.9996377	1.0255848	0.9862585	0.9898501	0.9945160	1.0012809	0.9926205
1.0171993	0.9905718	0.9793123	1.0033435	0.9896216	1.0057380	1.0403868	0.9848493	1.0050233
1.0020764	0.9845090	1.0143431	1.0083949	0.9987872	0.9893704	0.9985842	1.0003293	1.0293722
1.0235744	1.0022448	1.0026420	0.9852454	0.9950316	0.9911739	0.9954857	0.9666974	0.9899235
0.9885077	0.9941717	1.0346857	1.0279386	1.0330895	1.0254239	1.0126299	1.0231046	1.0022947
0.9971806	0.9539593	0.9485851	0.9400578	0.9732958	1.0279477	1.0502371	1.0808765	0.9873961
0.9995607	0.9820266	0.9422311	0.9842542	0.9522537	0.9802574	0.9763666	1.0195626	0.9986232
1.0391470	1.0549614	0.9992076	1.0031069	1.0152521	1.0016865	1.0057023	0.9607745	0.8956659
1.0082737	1.0191569	1.0138678	NA	NA	NA	NA	NA	NA

Random component of time series



Time series = Seasonal * Trend * Random

NA NA NA NA NA NA 148 148 136 119 104 118 115 126 141 135 125 149 170 170 158 133 114 140 145 150 178 163 172 178 199 199 184 162 146 166 171 180 193 181 183 218 230 242 209 191 172 194 196 196 236 235 229 243 264 272 237 211 180 201 204 188 235 227 234 264 302 293 259 229 203 229 242 233 267 269 270 315 364 347 312 274 237 278 284 277 317 313 318 374 413 405 355 306 271 306 315 301 356 348 355 422 465 467 404 347 305 336 340 318 362 348 363 435 491 505 404 359 310 337 360 342 406 396 420 472 548 559 463 407 362 405 417 391 419 461 472 535 NA NA NA NA NA NA


o
100
200
300
400
500
1950
1952
1954
1956
1958
1960
Time

3. Time Series Forecasting

Steps in Time Series Forecasting

1. Preliminary analysis.
2. Choosing and fitting models.
3. Using and evaluating a forecasting model.

Step#1:- Preliminary analysis.

1. If there is a relation between values of time series , then only that time series can be used for forecasting.

- Tools to check that relation are
  1. Autocorrelation (ACF)
  2. Correlogram
  3. Ljung–Box test

2. Stationary vs. Non stationary time series

1. Autocorrelation

- An autocorrelation is a **correlation** of the values of a variable with values of the same variable **lagged** one or more periods back.
- Autocorrelation refers to the correlation of a time series with its own past and future values.

Time	X	X Lag 1	X Lag 2
1	360		
2	350	360	
3	225	350	360
4	211	225	350
5	210	211	225
6	85	210	211
7	70	85	210
8	69	70	85
9	16	69	70
10	15	16	69

Autocorrelation (cont…)

Degree of autocorrelation (r):-

1. It is a measure of the internal correlation within a time series.
2. It assigning a value of +1 to strong positive association
3. -1 to strong negative association and
4. 0 to no association

$$r_k = \frac{\sum_{i=1}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Time	X	X Lag 1
1	360	
2	350	360
3	225	350
4	211	225
5	210	211
6	85	210
7	70	85
8	69	70
9	16	69
10	15	16

$$r_k = \frac{\sum_{i=1}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

r= 0.9

Time	X	X Lag 1
1	10	
2	-11	10
3	9	-11
4	-8	9
5	7	-8
6	-10	7
7	12	-10
8	-7	12
9	8	-7
10	-5	8

$$r_k = \frac{\sum_{i=1}^{N-k}(x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

r= - 0.8

Time	X	X Lag 1
1	1	
2	100	1
3	9	100
4	58	9
5	2	58
6	-10	2
7	2	-10
8	3	2
9	150	3
10	50	150

$$r_k = \frac{\sum_{i=1}^{N-k}(x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

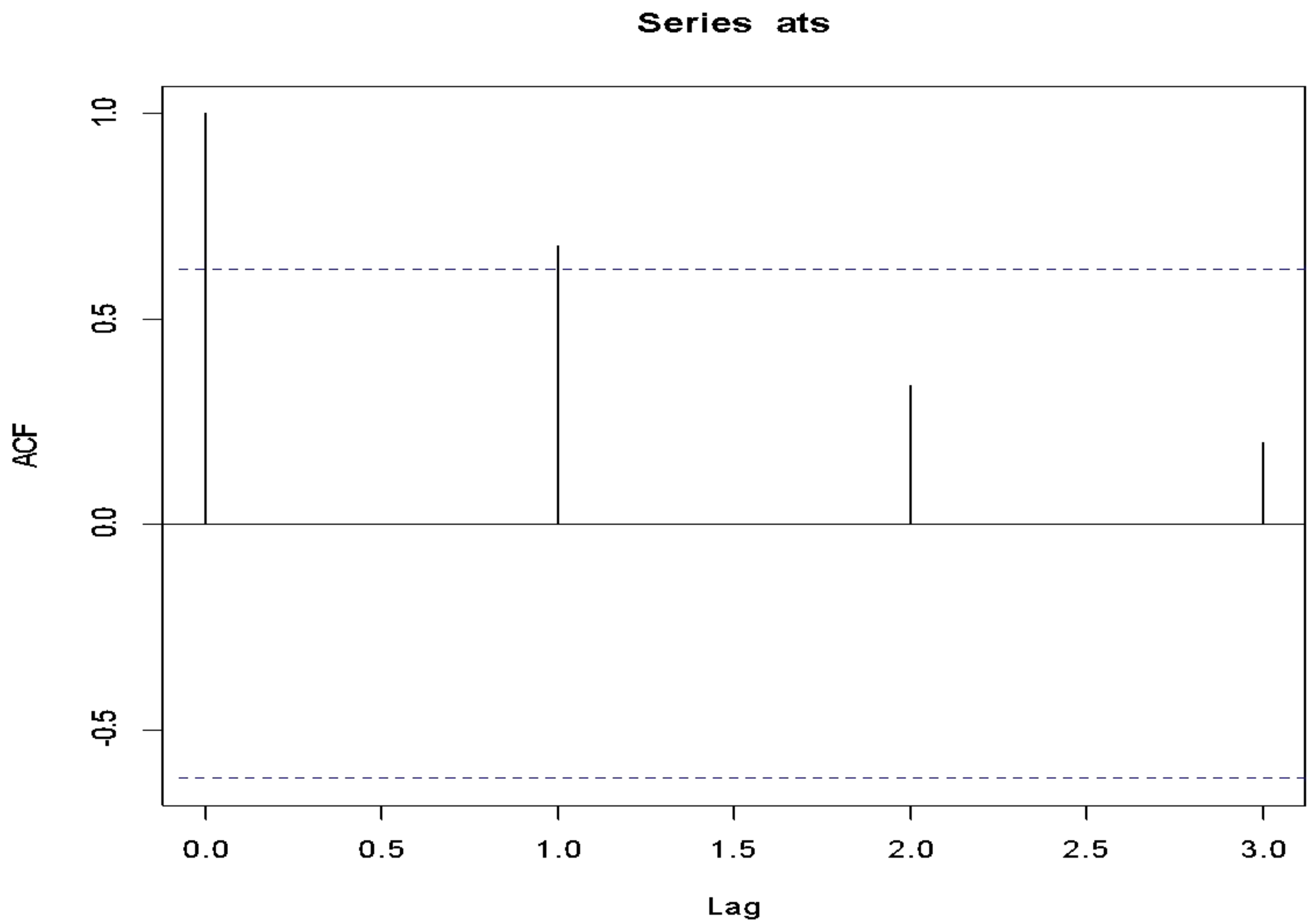
$$r = 0.04 \cong 0$$

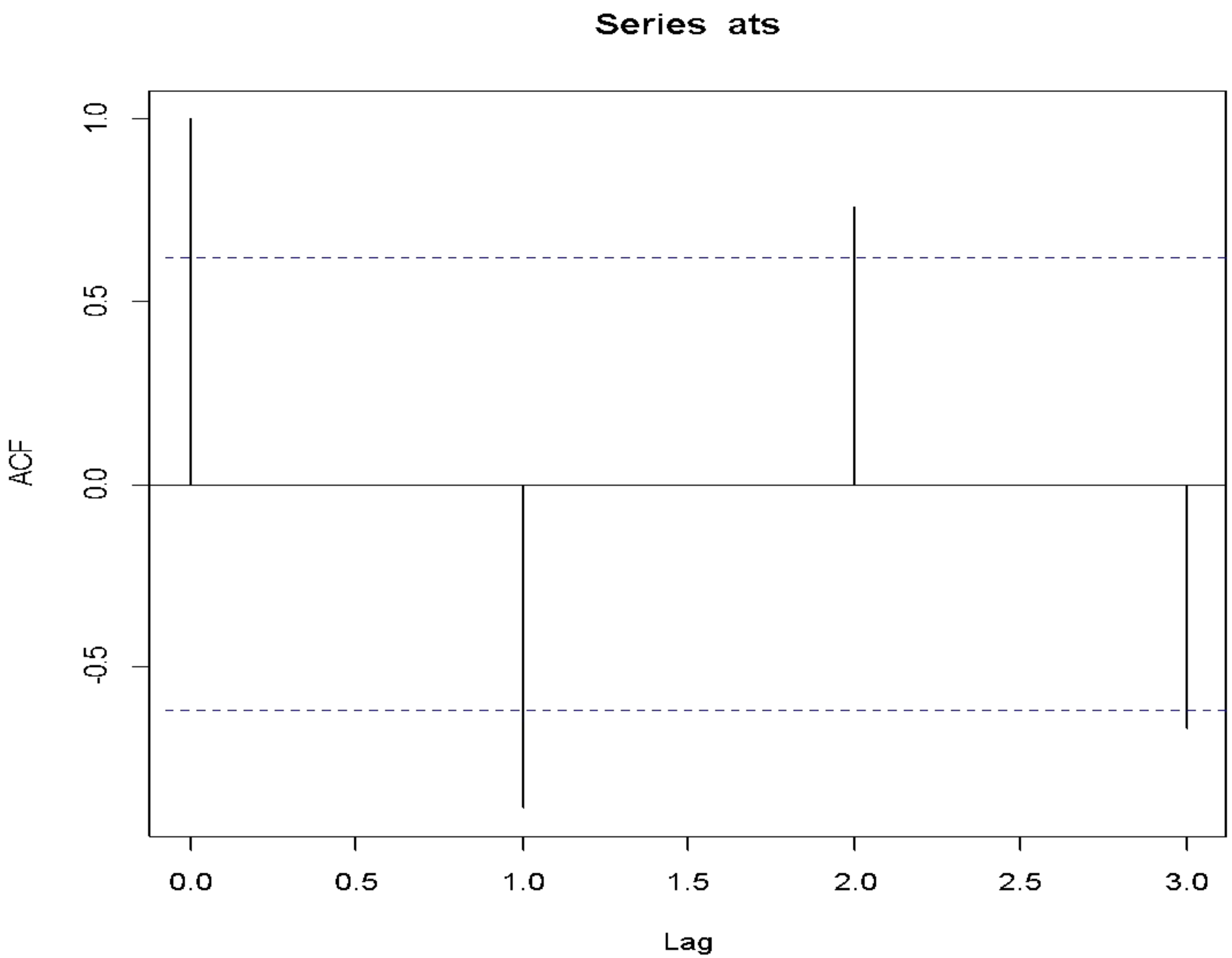
2. Correlogram

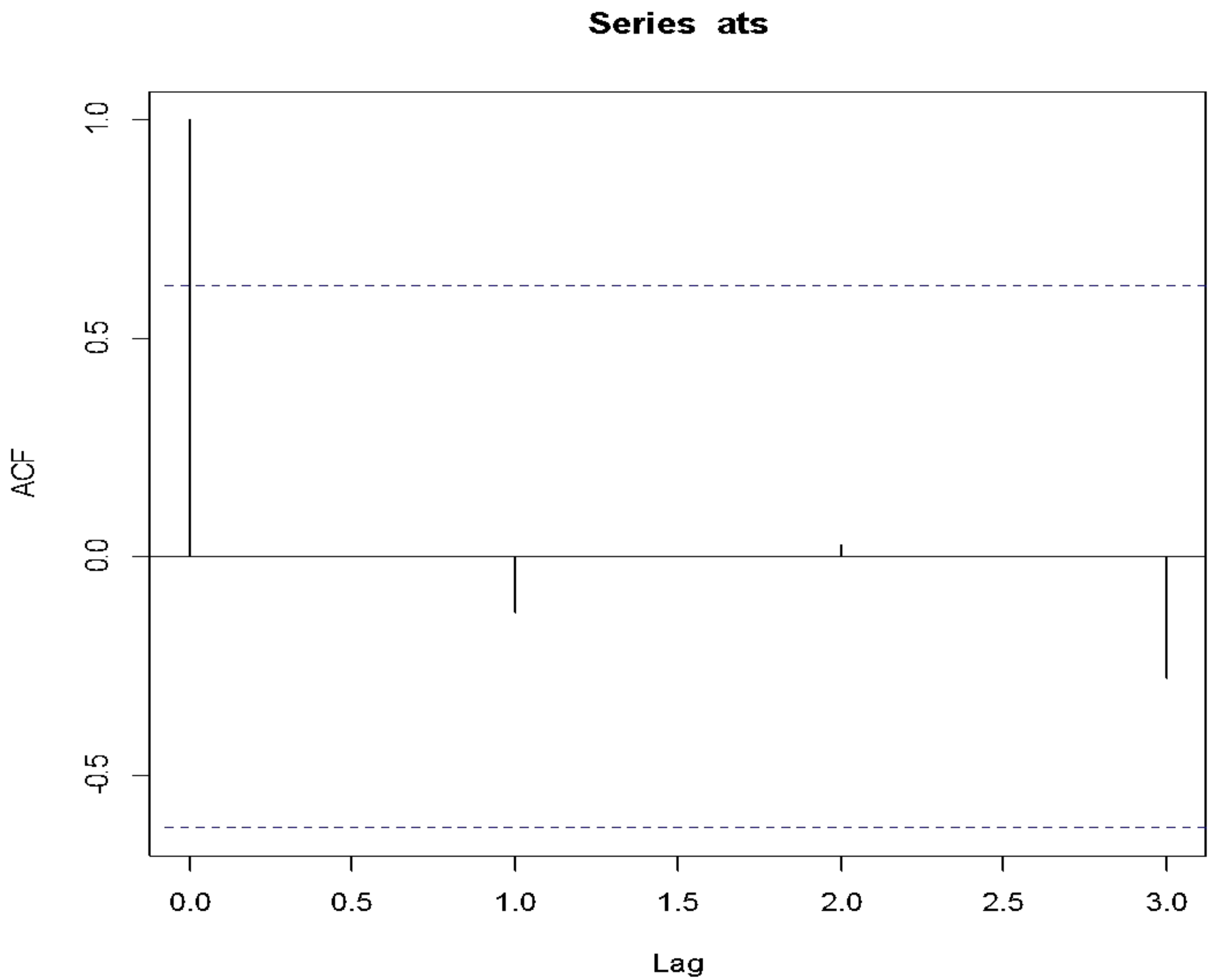
- Correlogram is plot of **lag(k)** autocorrelation versus **k**.
- i.e. the plot of the ACF against k.

- ACF is Autocorrelation function.

- As the ACF lies between -1 and +1, the correlogram also lies between these values.


Series ats
ACF
1.0
0.5
0.0
-0.5
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Lag


Series ats
ACF
1.0
0.5
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Lag


Series ats
1.0
0.5
0.0
-0.5
ACF
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Lag

3. Ljung-Box test

1. The **Ljung–Box test** named for Greta M. Ljung and George E. P. Box.

2. The autocorrelation function (ACF) is useful qualitative tools to assess the presence of autocorrelation at **individual lags**.

3. The Ljung-Box test is a more quantitative way to test for autocorrelation at **multiple lags** jointly.

4. Instead of testing randomness at each distinct lag, it tests the "overall" randomness based on a number of lags.

LJUNG-BOX TEST (CONT…)

5. This test is based on the statistic

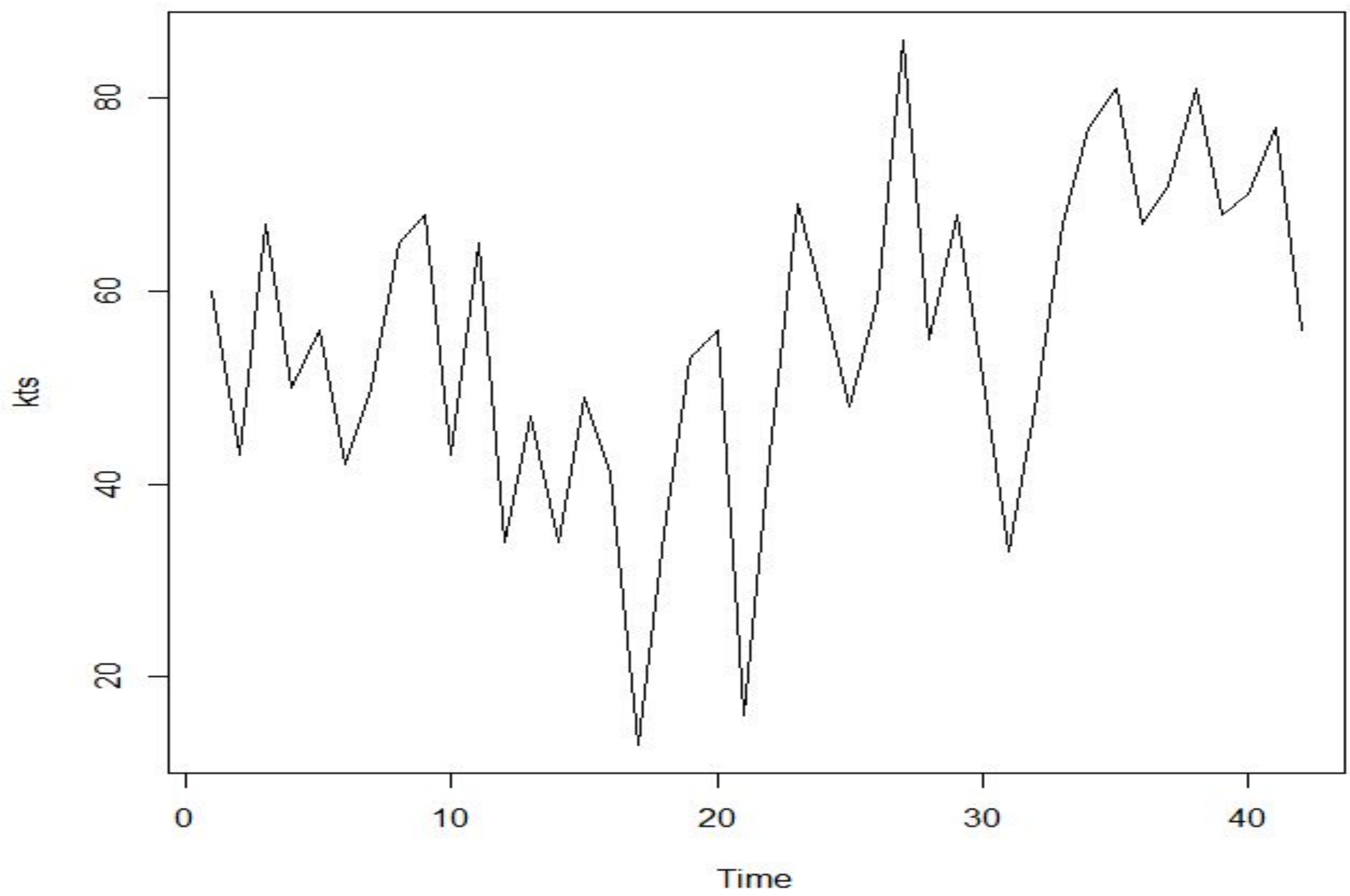
$$\chi^2(m) \approx Q(m) = n(n+2)\sum_{k=1}^{m}\left(\frac{r_k^2}{n-k}\right)$$

Where
1. **n** is the length of the time series.
2. m is the number of lags to test.
3. **r(k)** is autocorrelation coefficient at lag k

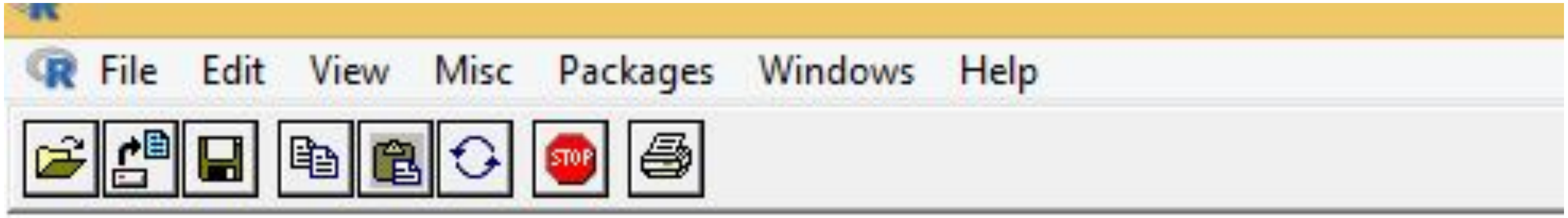
6. Large values of Q indicate that there are significant autocorrelations in the Time series.

7. When the number of observations is large, then the Q statistic has a **Chi-square distribution** with m-p -q **degrees of freedom**

60	13	67
43	35	77
67	53	81
50	56	67
56	16	71
42	43	81
50	69	68
65	59	70
68	48	77
43	59	56
65	86	
34	55	
47	68	
34	51	
49	33	
41	49	



kts
80
60
40
20
0
10
20
30
40
Time




```
> Box.test(kts,lag=20,type="Ljung-Box")

        Box-Ljung test

data:  kts
X-squared = 36.952, df = 20, p-value = 0.01186

> 
```

df = Degree of freedom

Hypothesis Testing

Hypothesis Testing

- Is also called *significance testing*
- Tests a claim about a parameter of population using evidence (data in a sample)
- The procedure is broken into steps

Hypothesis Testing Steps

A. Null and alternative hypotheses
B. Test statistic
C. P-value and interpretation

A. Null Hypothesis and Alternative Hypothesis

- The Null hypothesis assumes that any kind of **difference or significance** we see in a set of data is due to **chance**.

- The **alternative hypothesis** is the hypothesis that is accepted if the null hypothesis is rejected.

A. Null and Alternative Hypotheses

- Convert the research question to null and alternative hypotheses
- The **null hypothesis (H_0)** is a claim of "no difference in the population"
- The **alternative hypothesis (H_a or H_1)** claims "H_0 is false"
- Collect data and seek evidence against H_0 as a way of proving H_a (deduction)

A. Null and Alternative Hypotheses for Ljung-Box test

- In general, the Box-Ljung test is defined as:

- $\boldsymbol{H_0}$: There is **no** Autocorrelation in the time series.

- $\boldsymbol{H_a}$: There is Autocorrelation in the time series.

B. Test Statistic for Ljung-Box test

- Given a time series Y of length n, the test statistic is defined as:

$$\chi^2(m) \approx Q(m) = n(n+2)\sum_{k=1}^{m}\left(\frac{r_k^2}{n-k}\right)$$

- When the number of observations is large, then the Q statistic has a Chi-square distribution with
  m-p -q degrees of freedom

C. *P*-VALUE

- Convert **Chi-square distribution** to *P*-value :

	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32.000	34.267	37.146	39.252
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.790
18	6.265	8.231	22.760	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312
19	6.844	8.907	23.900	27.204	30.144	32.852	33.687	36.191	38.582	41.610	43.820
20	7.434	9.591	25.038	28.412	31.410	34.170	35.020	37.566	39.997	43.072	45.315

```
Box-Ljung test

data:  kts
X-squared = 36.952, df = 20, p-value = 0.01186
```

	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32.000	34.267	37.146	39.252
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.790
18	6.265	8.231	22.760	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312
19	6.844	8.907	23.900	27.204	30.144	32.852	33.687	36.191	38.582	41.610	43.820
20	7.434	9.591	25.038	28.412	31.410	34.170	35.020	37.566	39.997	43.072	45.315

C. P-VALUE

- *P*-value is the probability of obtaining a **test statistic** at least as extreme as the one that was actually observed, assuming that the **null hypothesis** is true.

- *P*-value tells us how likely it is that the result reported in a study is true and did not just occur because of chance.

C. Interpretation P-values

- Thus, smaller and smaller P-values provide stronger and stronger evidence against H_0

- Small P-value $\Rightarrow$ strong evidence against H_0

C. Interpretation

$P > 0.10 \Rightarrow$ non-significant evidence against H_0

$0.05 < P \leq 0.10 \Rightarrow$ marginally significant evidence

$0.01 < P \leq 0.05 \Rightarrow$ significant evidence against H_0

$P \leq 0.01 \Rightarrow$ highly significant evidence against H_0

```
        Box-Ljung test

data:  kts
X-squared = 36.952, df = 20, p-value = 0.01186
```

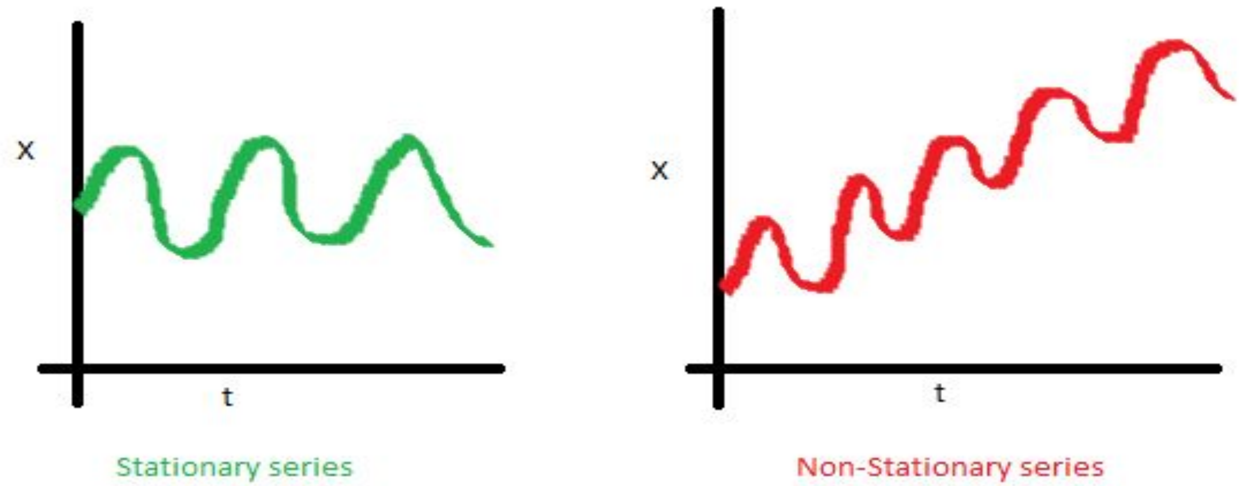
2. Stationary an non-stationary time series

- If time series is not stationary then we cannot build a time series model.
- If time series is non-stationary then first requisite becomes to stationarize the time series.
- And then try models to predict this time series
- Detrending, Differencing etc.

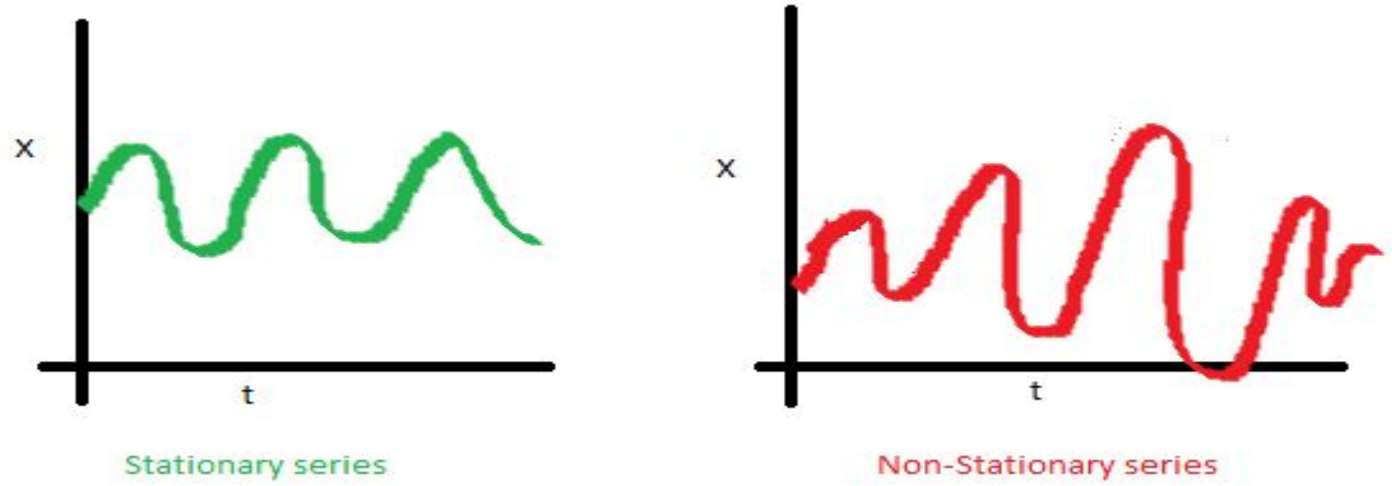
Stationary Series

There are three basic criterion for a series to be classified as stationary series :

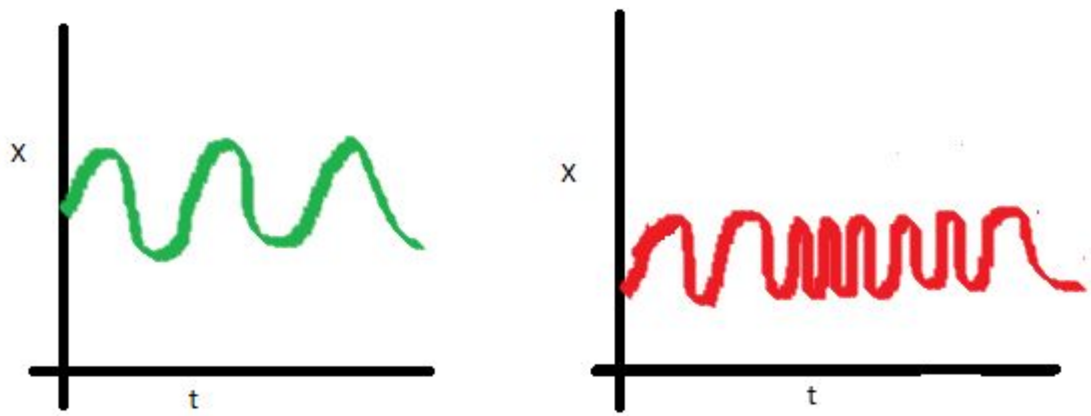
1. The **mean** of the series should not be a function of time rather should be a constant.



2. The **variance** of the series should not a be a function of time.



3. The **covariance** of the i th term and the (i + m) th term should not be a function of time.



Stationary series

Non-Stationary series

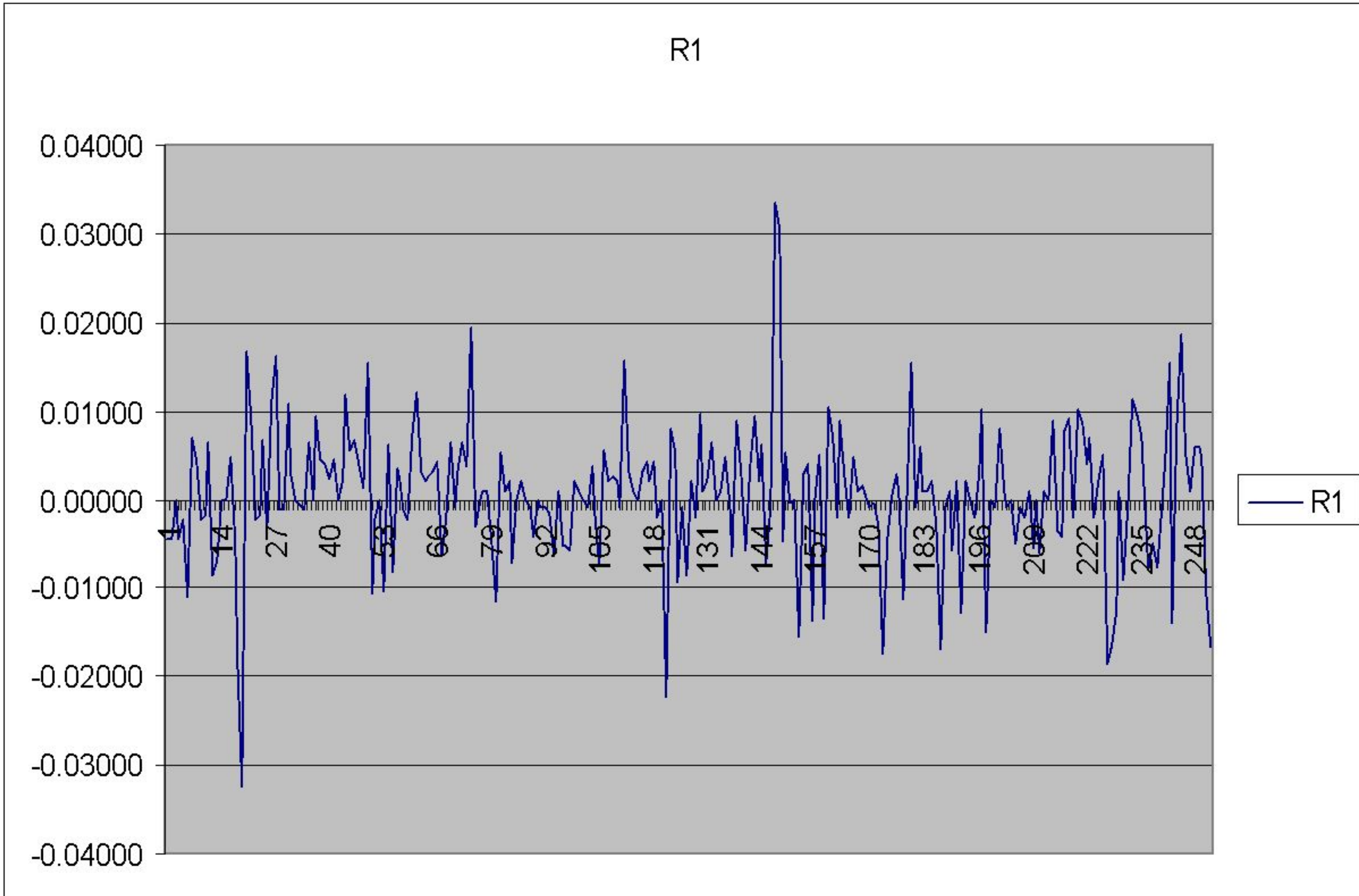
Stationary Time Series

- In stationary time series joint probability of a series doesn't change over time. i.e. mean and variance remains constant over time.

- F(Yt) = F(Yt + k), where F is joint probability distribution.

- No trend in the series.

Example of stationary time series



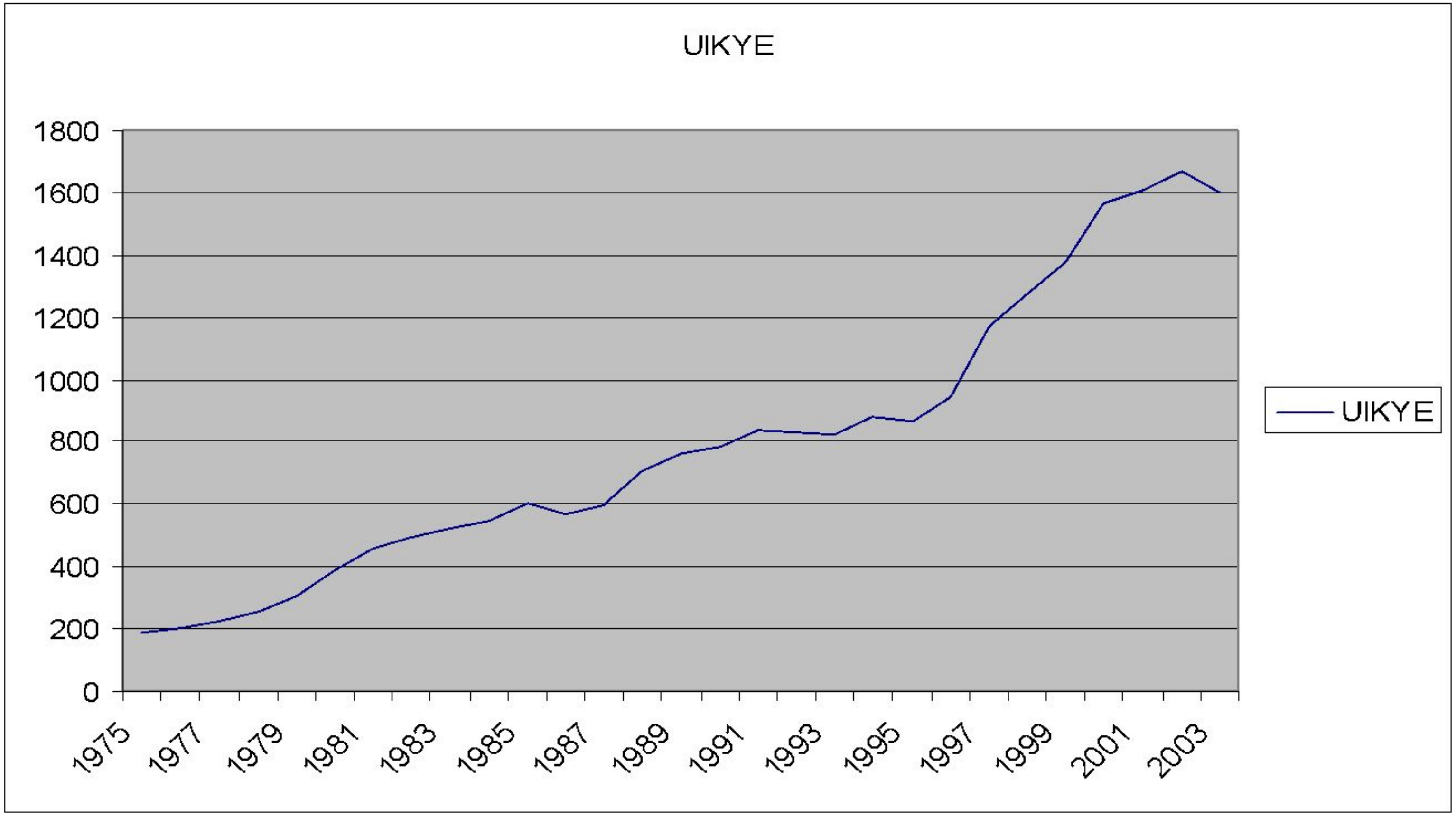
Non-Stationary time series

- In Non-stationary time series joint probability of a series change over time. i.e. mean and variance doesn't remains constant over time.

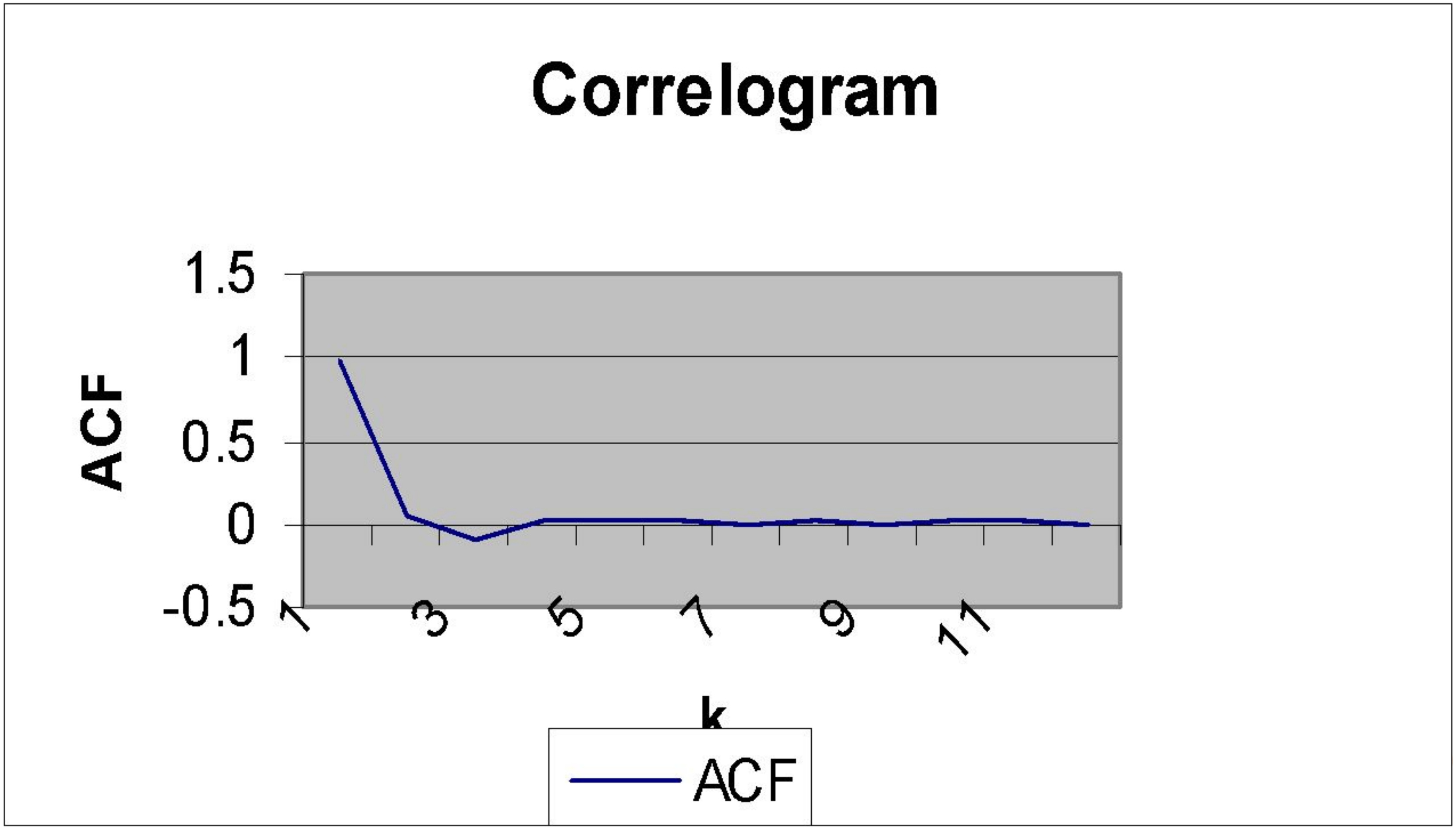
- F(Yt) != F(Yt + k), where F is joint probability distribution.

- There is trend in the series.

Non-stationary Series



Stationary time series



- Correlogram can be used to determine stationarity, if the ACF falls immediately from 1 to 0, then equals about 0 thereafter, the series is stationary.
- If the ACF declines gradually from 1 to 0 over a prolonged period of time, then it is not stationary.

Step#2:- Choosing and fitting models

Time series models for forecasting

- AutoRegressive Model (AR)
- Moving Average Model (MA)
- AutoRegressive Moving Average Model (ARMA)
- Autoregressive Integrated Moving Average Model (ARIMA)

AutoRegressive Model: (AR)

1. An autoregressive (AR) model **predicts future behaviour based on past behaviour.**

2. It's used for forecasting when there is some correlation between values in a time series and the values that precede and succeed them.
3. The process is basically a **linear regression** of the data in the current series against one or more past values in the same series.

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p,} \varepsilon_t)$$

The most often seen form of the equation is a linear form AR(p):

$$y_t = b_0 + \sum_{i=1}^{p} b_i y_{t-i} + e_t$$

where:

y_t – the dependent variable values at the moment *t*,

y_{t-i} (i = 1, 2, ..., p) – the dependent variable values at the moment *t-i*,

bo, bi (i=1,..., p) – regression coefficient,

p – autoregression order,

e_t – error term.

If p=1 then called AR(1) model.

$$\boxed{y_t = b_0 + b_1 y_{t-1} + e_t}$$

If p=2 then called AR(2) model.

$$\boxed{y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + e_t}$$

Example

t	y_t
1	1.89
2	2.46
3	3.23
4	3.95
5	4.56
6	5.07
7	5.62
8	6.16
9	6.26
10	6.56
11	6.98
12	7.36
13	7.53
14	7.84
15	8.09

AR(2)

$$y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + e_t$$

AR(2)

$$y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + e_t$$

t	y_t	y_{t-1}	y_{t-2}
1	1.89	-	-
2	2.46	1.89	-
3	3.23	2.46	1.89
4	3.95	3.23	2.46
5	4.56	3.95	3.23
6	5.07	4.56	3.95
7	5.62	5.07	4.56
8	6.16	5.62	5.07
9	6.26	6.16	5.62
10	6.56	6.26	6.16
11	6.98	6.56	6.26
12	7.36	6.98	6.56
13	7.53	7.36	6.98
14	7.84	7.53	7.36
15	8.09	7.84	7.53

Moving Average Model: (MA)

- Rather than using **past values** of the forecast variable in a regression.

- Moving average model uses **past forecast errors** in a regression-like model.

$$y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots.. + \theta_q e_{t-q}$$

- Where c and $\theta_1, \ldots\ldots,$ θ_q are constants. and e(t) is error term

Moving Average Model: (MA)

$$y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots.. + \theta_q e_{t-q}$$

- We refer to this as an **MA(q) model**.
- As we do not *observe* the values of e(t), so it is not really regression in the usual sense
- So each value of y(t) can be thought of as a **weighted moving average of the past few forecast errors**.
- This is different than moving average *smoothing*

AutoRegressive Moving Average Model: (ARMA)

1. There are situations where the time-series may be represented as a mix of both **AR** and **MA** models referred as **ARMA(p, q)** model

2. The general form of such a time-series model, which depends on **p** of its own past values and **q** past values of white noise(error) disturbances, takes the form:

3. $Y(t) = \beta 0 + \beta 1\ Y(t-1) + \beta 2\ Y(t-2) + \ldots\ldots + \beta p\ Y(t-p) + \varepsilon(t) + \varphi 1\ \varepsilon(t-1) + \varphi 2 \varepsilon(t-2) + \ldots + \varphi p\ \varepsilon(t-q)$

Autoregressive Integrated Moving Average Model : **(ARIMA)**

1. MA, AR and ARMA models can not handle non-stationary time-series.

2. A time-series which is non-stationary (i.e. the series that has trend) can be made stationary after differentiating.

3. A series which is stationary after being differentiated once is said to be integrated of order 1 and is denoted by **I(1).**

4. In general **I(d)** where d is order of integration.

Autoregressive Integrated Moving Average Model : (ARIMA)

ARIMA (p ,q, r)

where p □ AR terms

q □ no. of differentiation

r □ no. of MA terms

ARIMA(1, 0, 0) □ AR model

ARIMA(0, 0, 1) □ MA model

ARIMA(1, 0, 1) □ ARMA model

Step#3:- Evaluating a forecasting model

- The errors in time series i.e. Actual value minus predicted value.
- we analyse these errors in Time Series for accuracy purpose.
- These error present in list called **RESIDUALS**.
- If there is no auto-correlation between RESIDUALS then forecast is pretty good.
- For this we use AUTOCORRELATION FUNCTION (ACF)

11. Case study (work out in R)

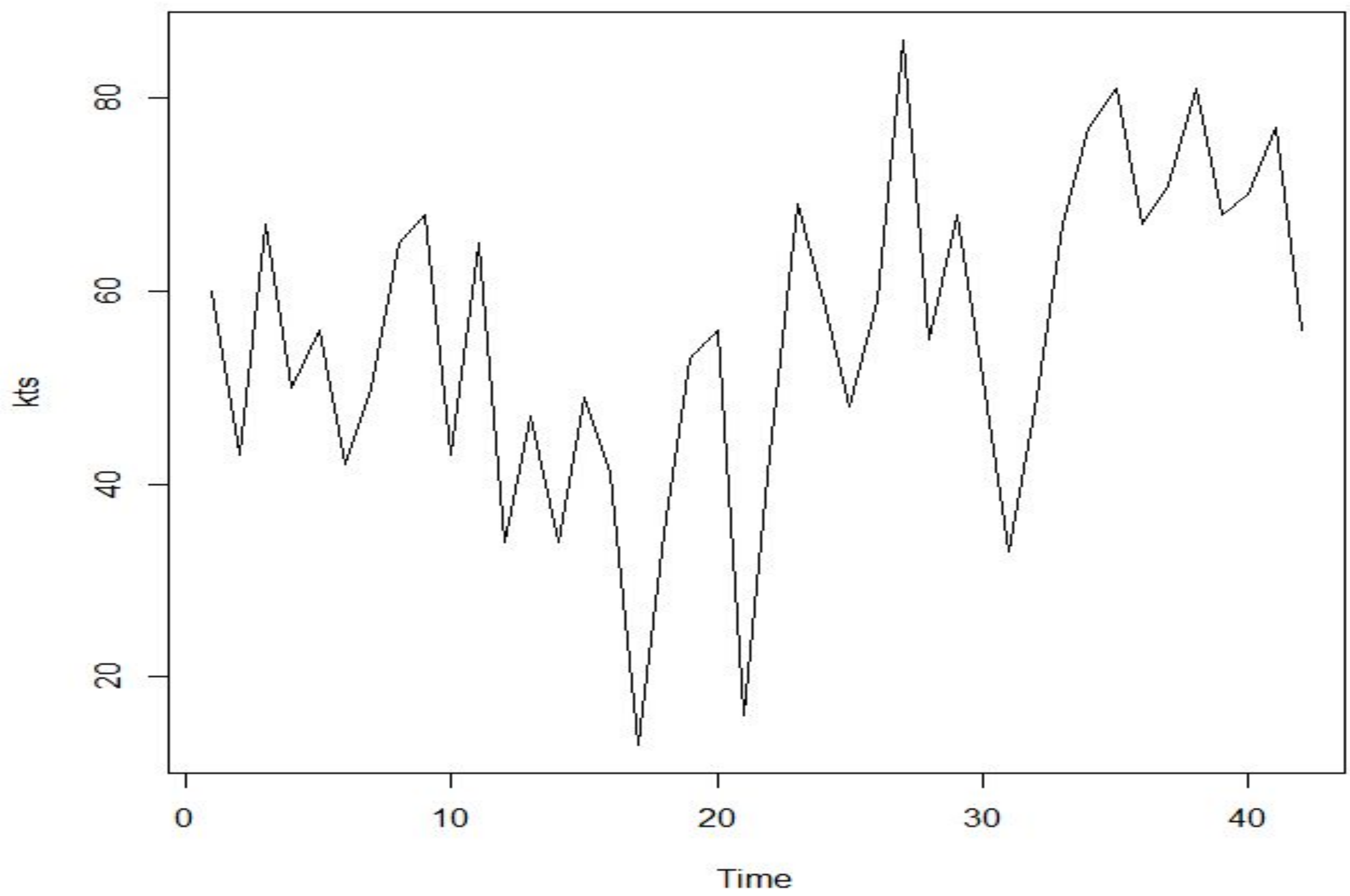
Dataset comprising of the age of death of 42 successive kings of England, starting with William the Conqueror (original source: Hipel and Mcleod, 1994).

The data set looks like this:

Age of Death of Successive Kings of England
#starting with William the Conqueror
#Source: McNeill, "Interactive Data Analysis"
60
43
11
67
50
56
42
50
65
.....

R Console

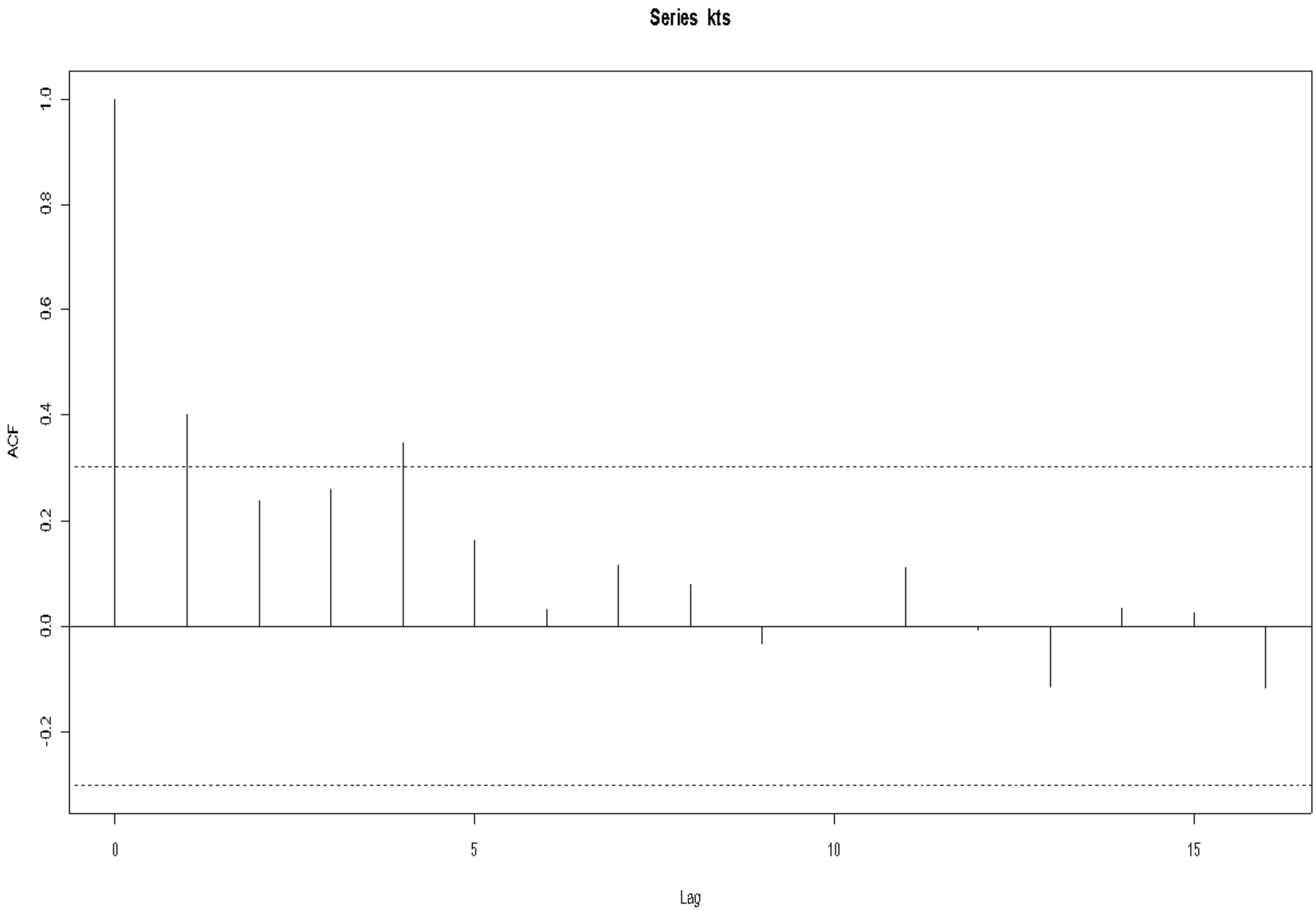
```
> # Importing data in R
>
> kings<-scan("kings.dat",skip=3)
Read 42 items
>
> # Converting data in Time Series
>
> kts<-ts(kings,start=c(1))
>
> plot.ts(kts)
>
```



kts
Time
20
40
60
80
0
10
20
30
40

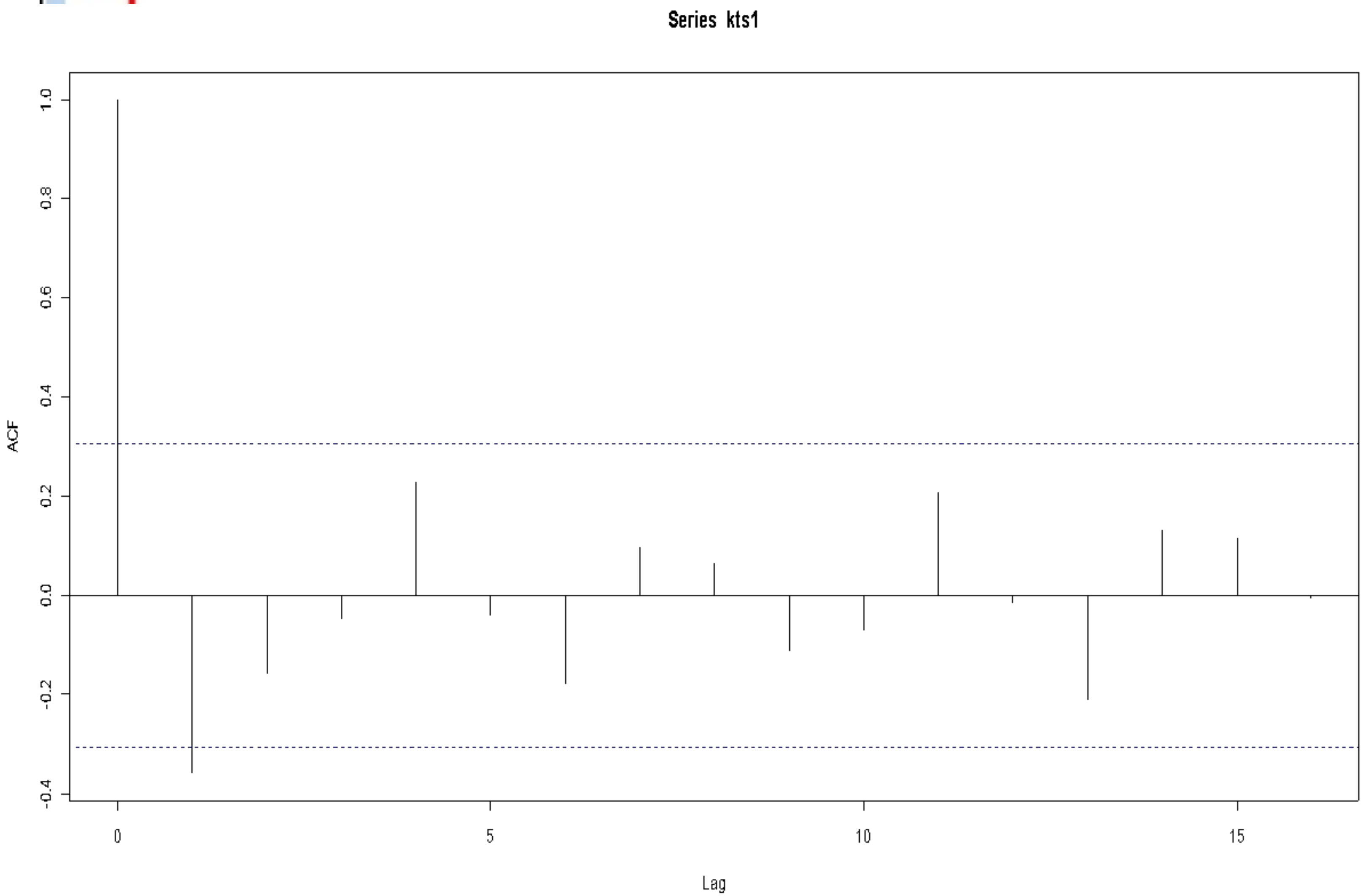
File Edit View Misc Packages Windows Help

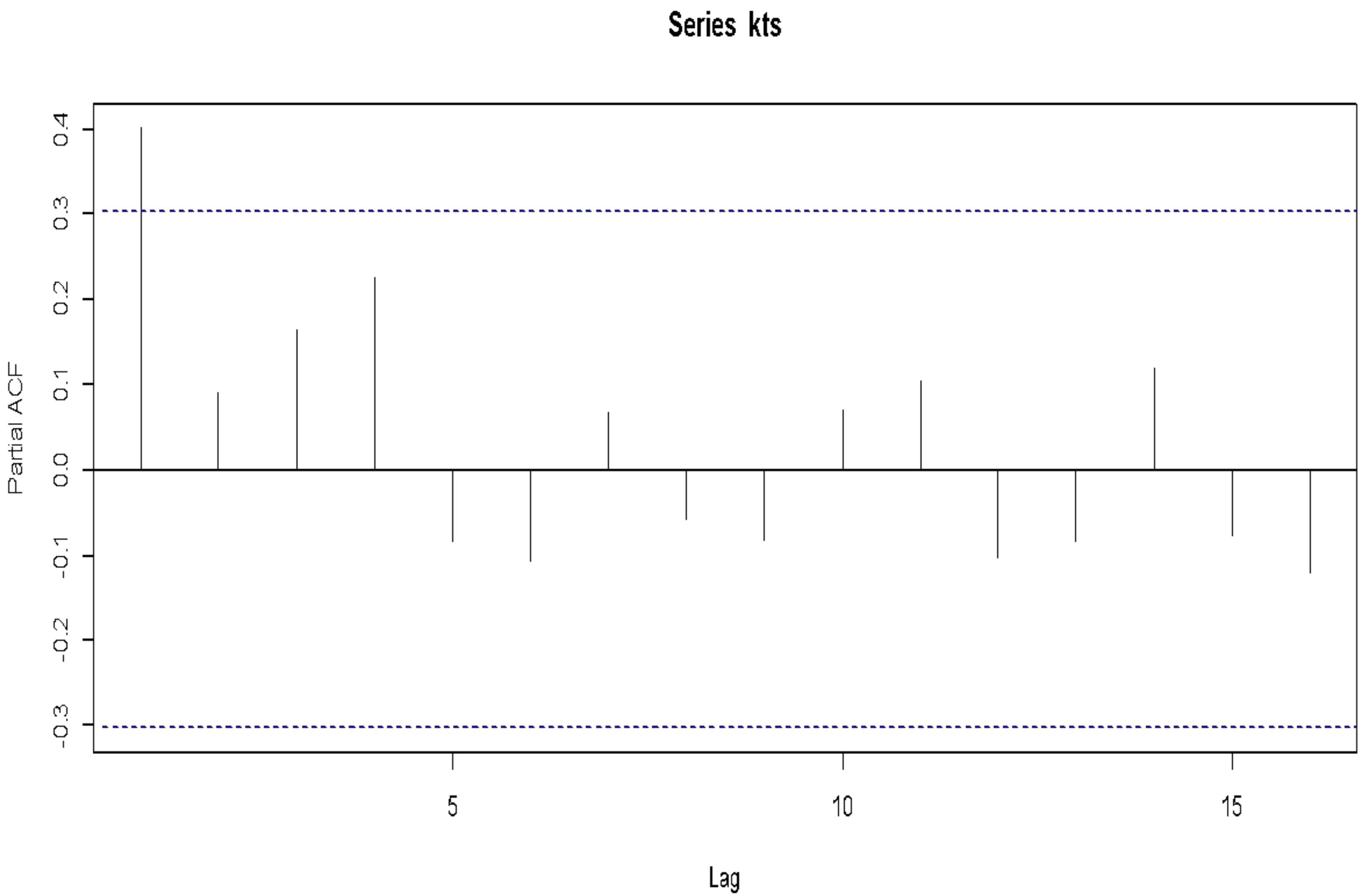
```
> Box.test(kts,lag=20,type="Ljung-Box")

        Box-Ljung test

data:  kts
X-squared = 36.952, df = 20, p-value = 0.01186

>
```



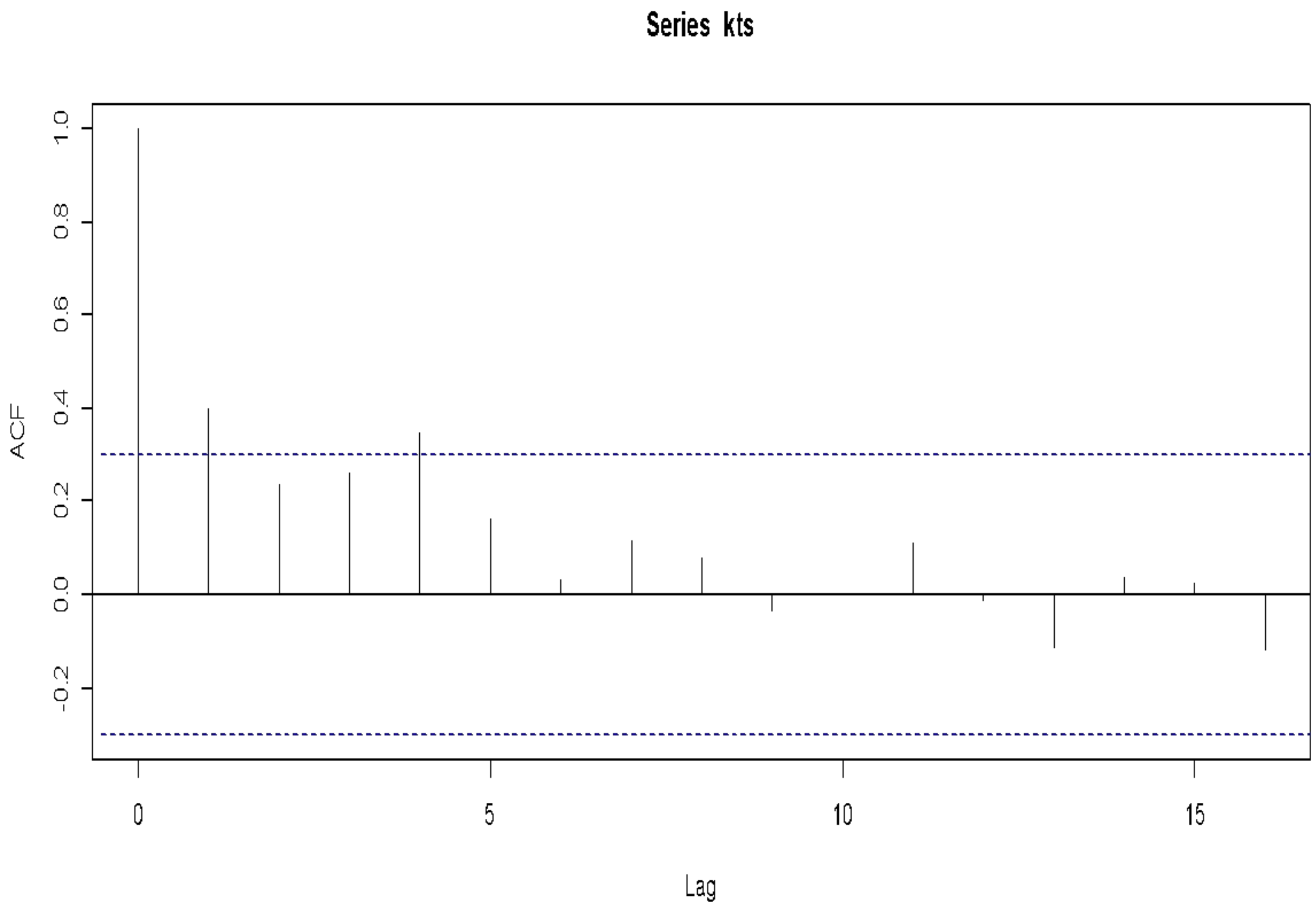
Series kts
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
Lag

```
> kts1<-diff(kts,difference=1)
> acf(kts1)
```





To know order of AR(p)
Here p=0



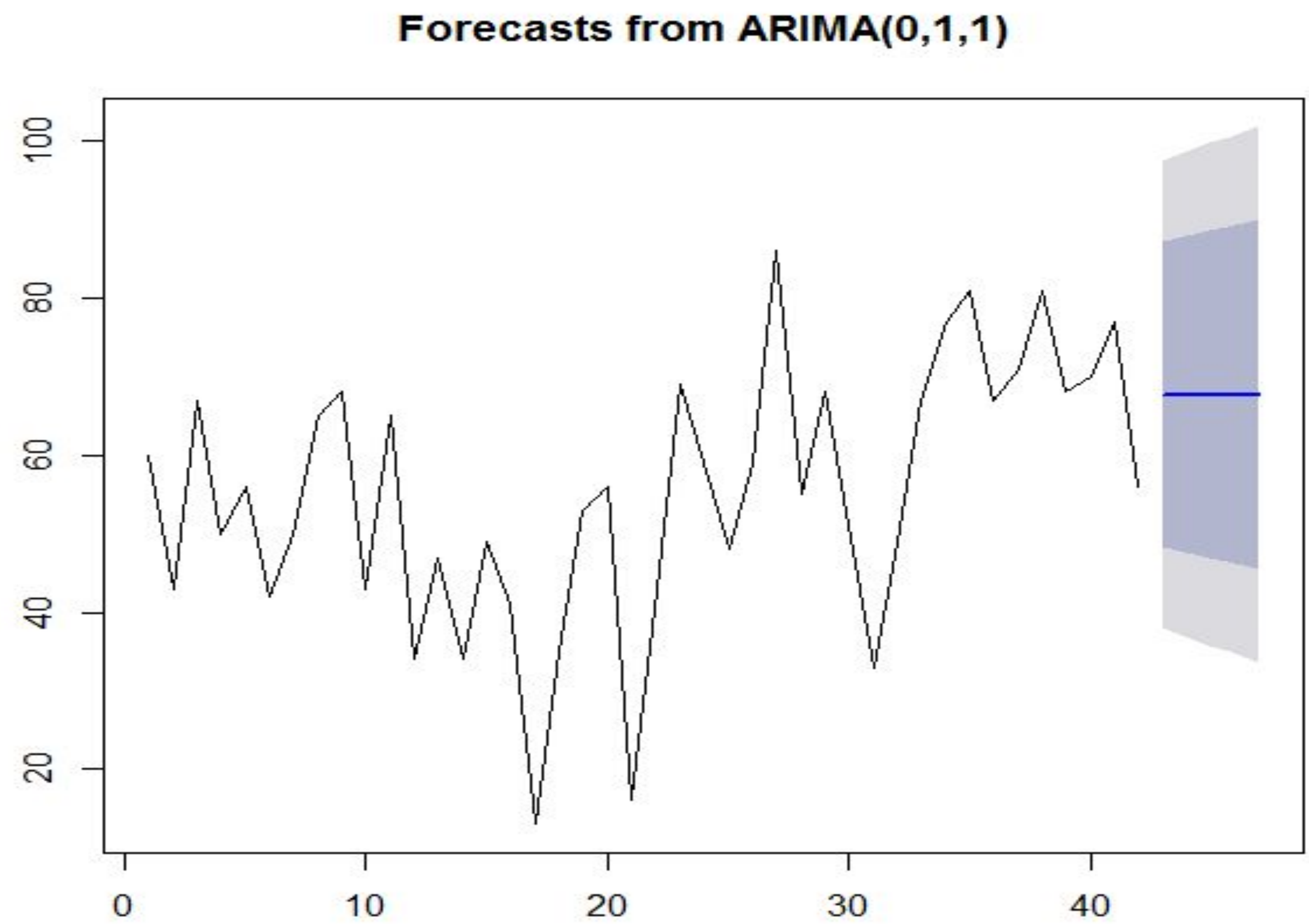
To know order of MA(q)
Here q=1

R Console

```
> # finding best possibe ARIMA model
> library(forecast)
> auto.arima(kts)
Series: kts
ARIMA(0,1,1)

Coefficients:
          ma1
      -0.7218
s.e.   0.1208

sigma^2 estimated as 236.2:  log likelihood=-170.06
AIC=344.13   AICc=344.44   BIC=347.56
> # finding best possibe ARIMA model
> ktsarima<- arima(kts,order=c(0,1,1))
> 
```

R Console

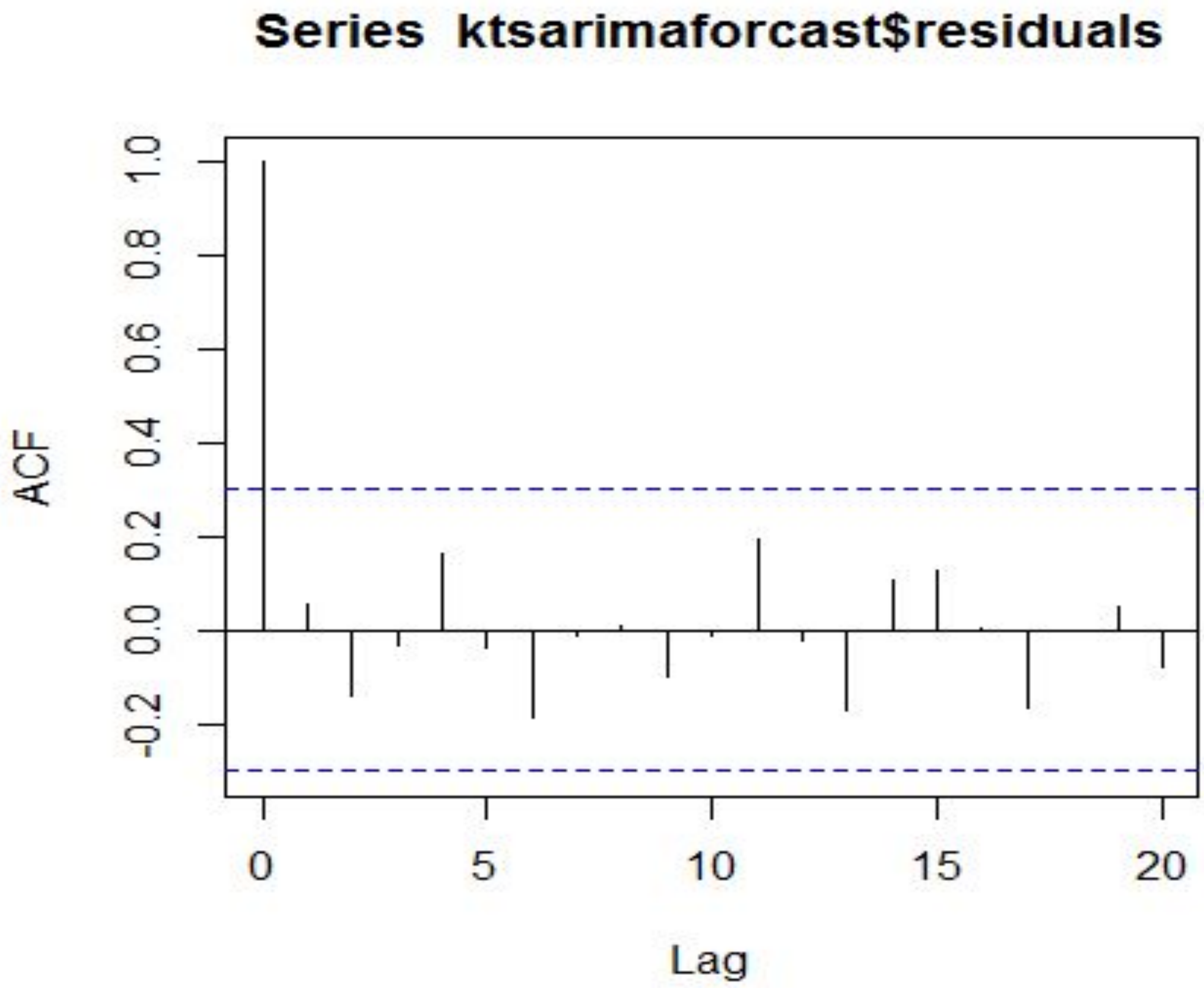
```
> ktsarimaforecast<-forecast(ktsarima,h=5)
> ktsarimaforecast
   Point Forecast    Lo 80    Hi 80    Lo 95     Hi 95
43       67.75063 48.29647 87.20479 37.99806  97.50319
44       67.75063 47.55748 87.94377 36.86788  98.63338
45       67.75063 46.84460 88.65665 35.77762  99.72363
46       67.75063 46.15524 89.34601 34.72333 100.77792
47       67.75063 45.48722 90.01404 33.70168 101.79958
>
> # Plotting Time Series forecast
>
> plot(ktsarimaforecast)
>
```



Forecasts from ARIMA(0,1,1)
100
80
60
40
20
0
10
20
30
40

R Console

```
> acf(ktsarimaforecast$residuals,lag.max=20)
> Box.test(ktsarimaforecast$residuals,lag=20,type="Ljung-Box")

        Box-Ljung test

data:  ktsarimaforecast$residuals
X-squared = 13.584, df = 20, p-value = 0.8509

>
```


Series ktsarimaforcast$residuals
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
20
Lag

THANK YOU