سری‌های زمانی

سری زمانی و تحلیل سری زمانی

- منظور از یک سری زمانی مجموعه‌ای از داده‌های آماری است که در فواصل زمانی مساوی و منظمی جمع‌آوری شده باشند.
- مثال: فروش ماهانه یک کارخانه طی سه سال گذشته
- مثال: میزان متوسط بارش سالیانه در کشور طی پنجاه سال گذشته
- تحلیل سری زمانی نیز یک از روش‌های تحلیل چنین داده‌هایی است.
- مثال: تشخیص روند تغییرات ارزش سهام با توجه به داده‌های جمع‌آوری شده در طول یک سال می‌تواند تحلیل سری زمانی نامیده شود.

سری زمانی و تحلیل سری زمانی

- هدف:

کشف ***الگو*** رفتار داده ها درگذشته و ***پیش بینی*** رفتار آینده با استفاده از داده‌های سری زمانی

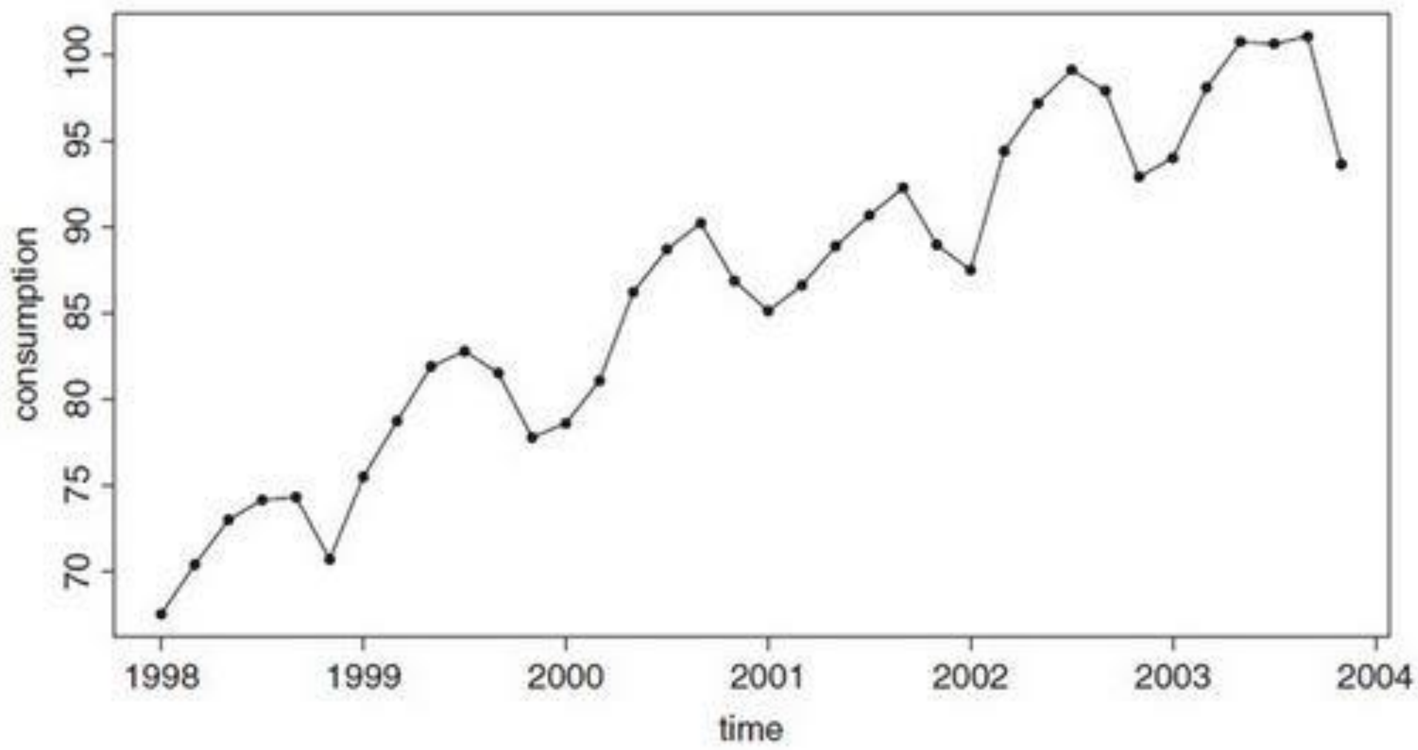
سری زمانی

- متغیر تصادفی: در آمار و احتمال، **متغیر تصادفی** متغیری است که مقدار آن از اندازه‌گیری یک فرایندهای تصادفی بدست می‌آید.
- اگر فرآیند تصادفی، میزان فروش کارخانه باشد و متغیر t نشاندهنده زمان، متغیر تصادفی X(t) میزان فروش کارخانه در ماه t را نشان می‌دهد.
- سری زمانی را می‌توان با متغیر تصادفی X(t) نشان داد:
- $X(t) \quad t = 1, 2, 3, \ \ldots$
- سری زمانی، برداری از داده‌ها که در فواصل زمانی مشخص دریافت و برحسب زمان مرتب شده‌اند.

سری زمانی

X = [22,71,74, 75, 75, 70, 76, 78, 82, 83, 82, … , 94]

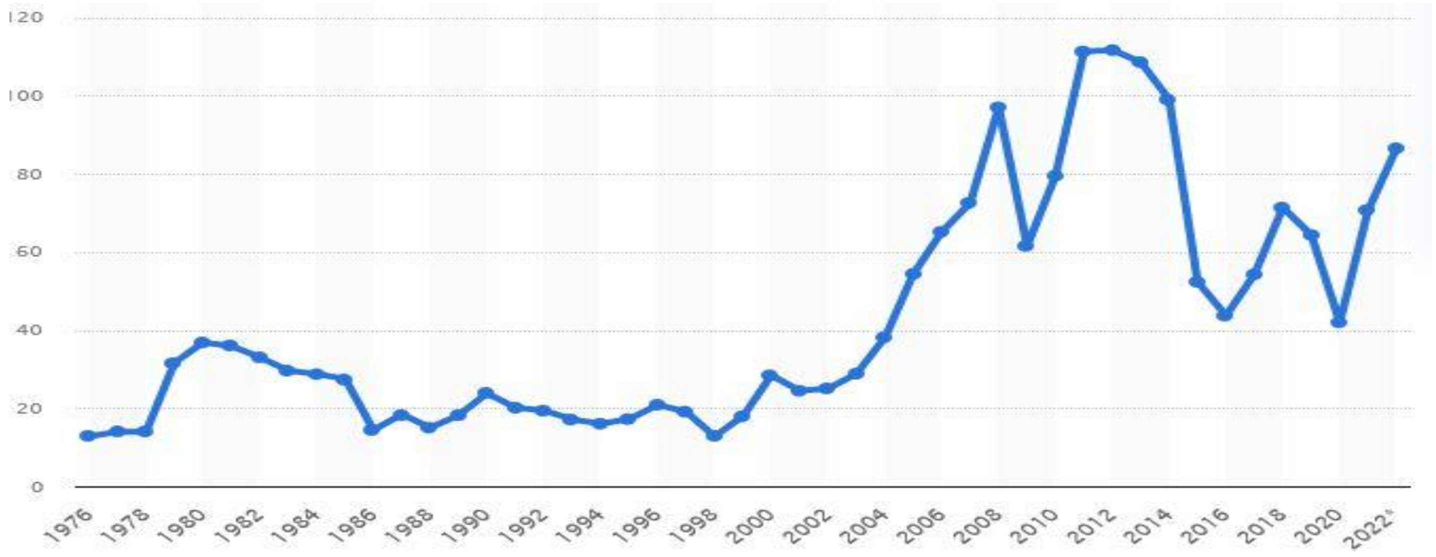
t = [1, 2, 3, …, 36]

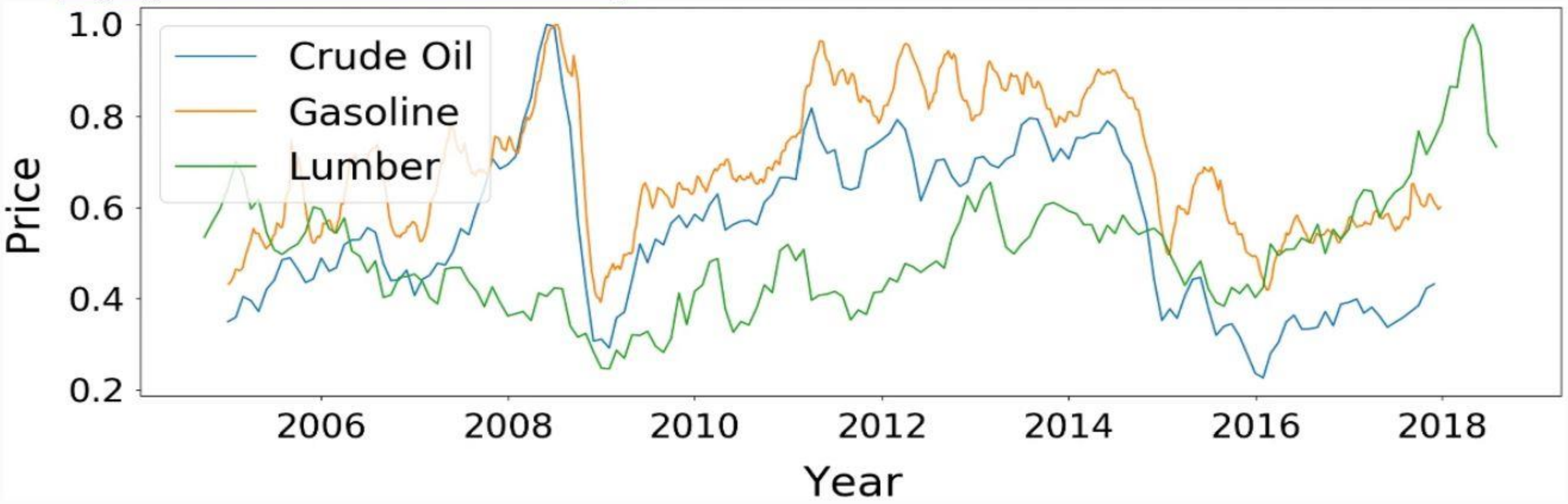


انواع متغیر سری زمانی

- تک متغیره: یک ویژگی از فرآیند تصادفی مورد بررسی قرار می گیرد. به طور مثال میزان فروش کلی یک کارخانه، قیمت نفت
- چند متغیره: بیشتر از یک ویژگی از فرآیند تصادفی مورد بررسی قرار می گیرد. به طور مثال میزان فروش هر یک از محصولات کارخانه، قیمت سوختهای فسیلی به تفکیک

انواع متغیر سری زمانی





انواع سری زمانی

- **گسسته:** برای سری زمانی $X(t)$، اگر نمونه‌برداری از فرآیند تصادفی به صورت گسسته انجام شودمثلا ساعتی، روزانه، ماهانه و ... باشد آنگاه سری زمانی گسسته است. (مثالهایی که دیدیم همگی گسسته بودند.)
- **پیوسته:** اگر t پیوسته باشد، آنگاه سری زمانی $X(t)$ نیز پیوسته خواهد بود. به طور مثال دمای هوا و یا دبی رودخانه

انواع سری زمانی

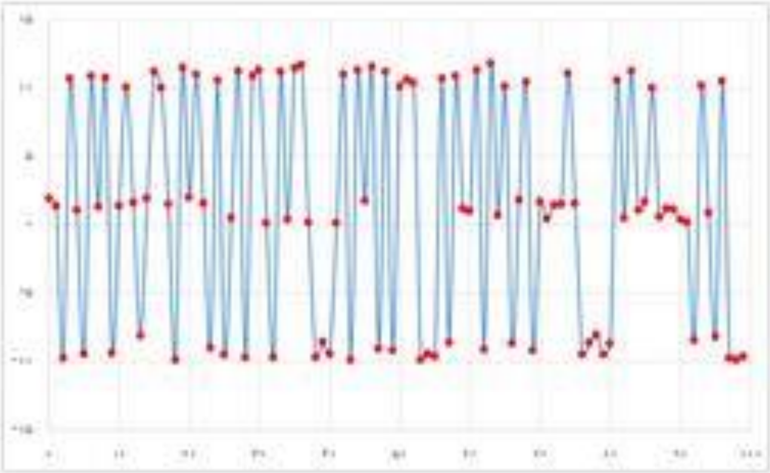
- **گسسته:** برای سری زمانی(X(t، اگر نمونه‌برداری از فرآیند تصادفی به صورت گسسته انجام شودمثلا ساعتی، روزانه، ماهانه و ... باشد آنگاه سری زمانی گسسته است. (مثالهایی که دیدیم همگی گسسته بودند.)
- **پیوسته:** اگر t پیوسته باشد، آنگاه سری زمانی X(t) نیز پیوسته خواهد بود. به طور مثال دمای هوا و یا دبی رودخانه

مولفه‌های سری زمانی

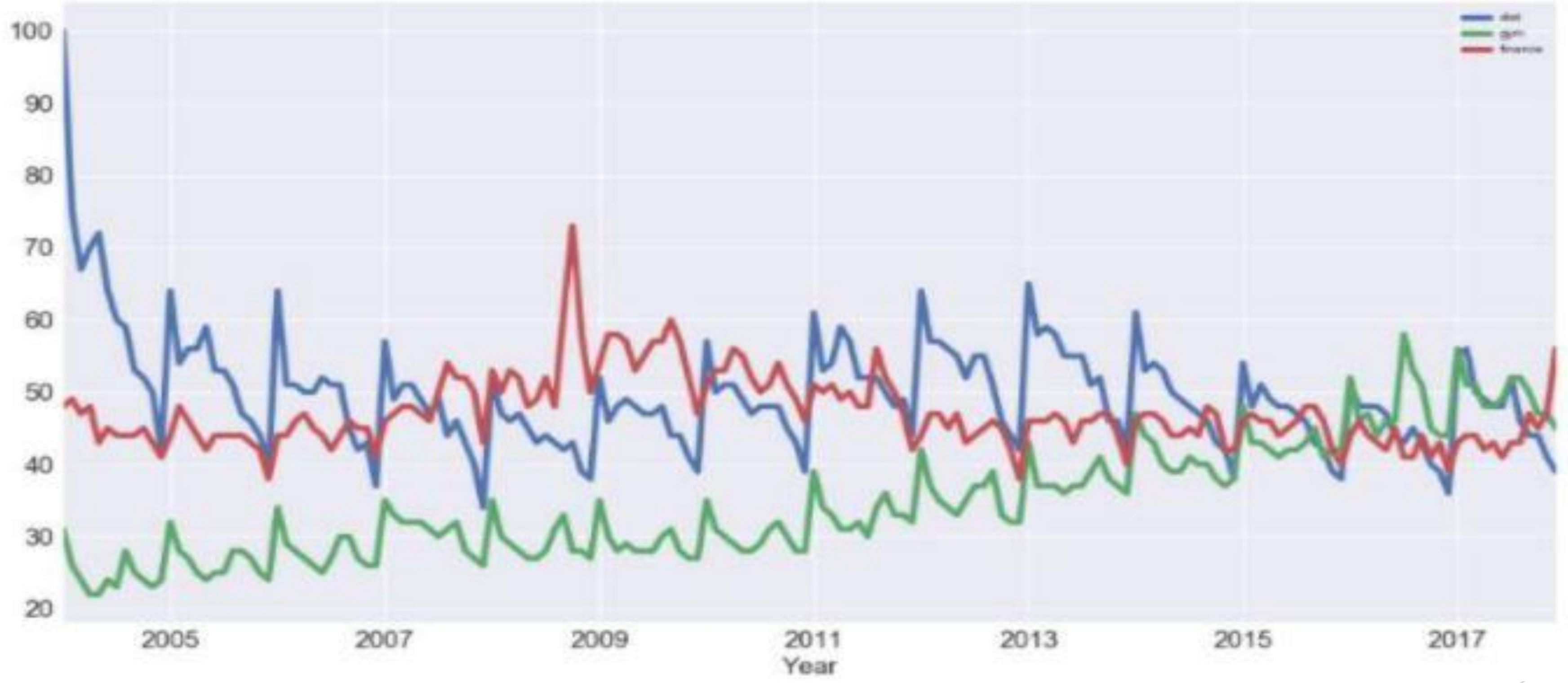
- چهار ویژگی الگوی رفتاری یک سری زمانی:
- روند (trend)
- فصلی (seasonal)
- غیر معمول (irregular)
- تناوبی (cyclic)

مولفه‌های سری زمانی—روند

- تمایل سری زمانی به افزایش، کاهش یا حتی ثابت بودن، را روند سری زمانی گویند.

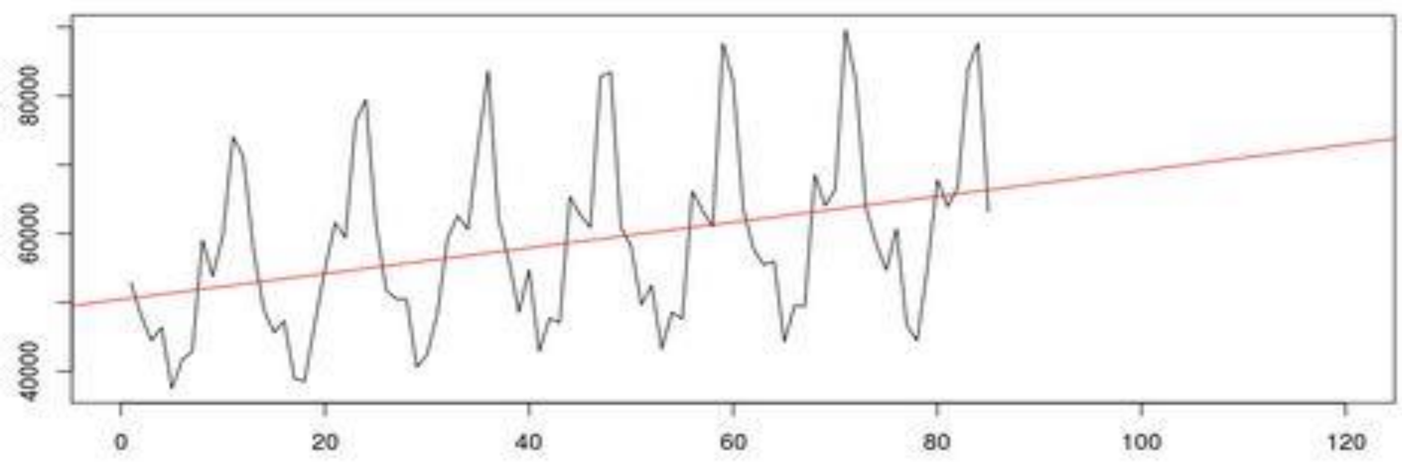
روند کاهشی	روند ثابت	روند افزایشی
		

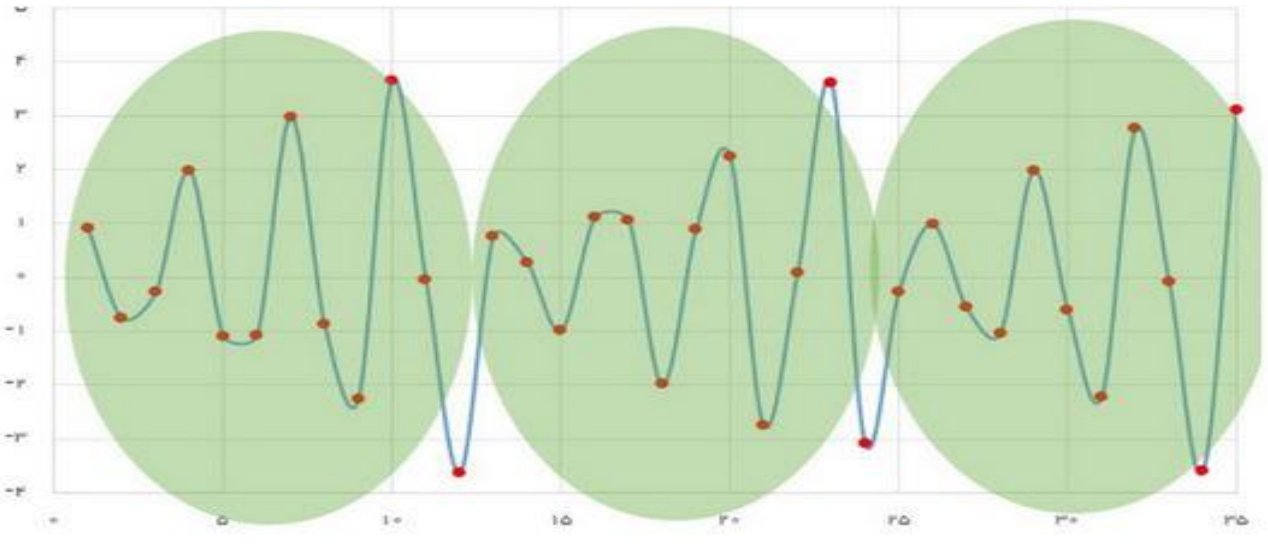
مولفه‌های سری زمانی–روند



مولفه‌های سری زمانی–فصلی

- فصلی: تغییراتی که در بازه‌های کوچک و با فواصل زمانی یکسان تکرار می‌شوند.

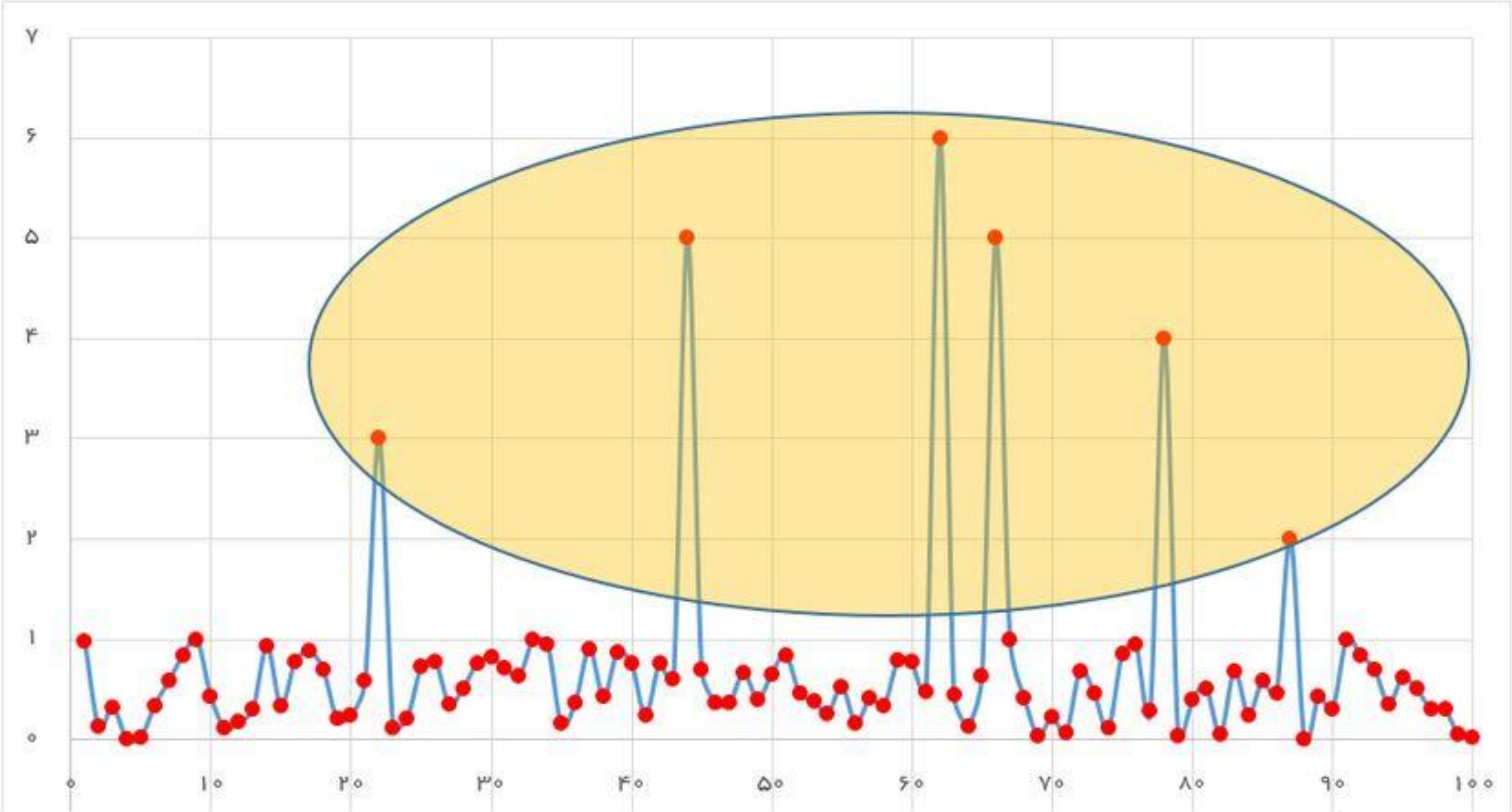




تغییرات فصلی در سری زمانی

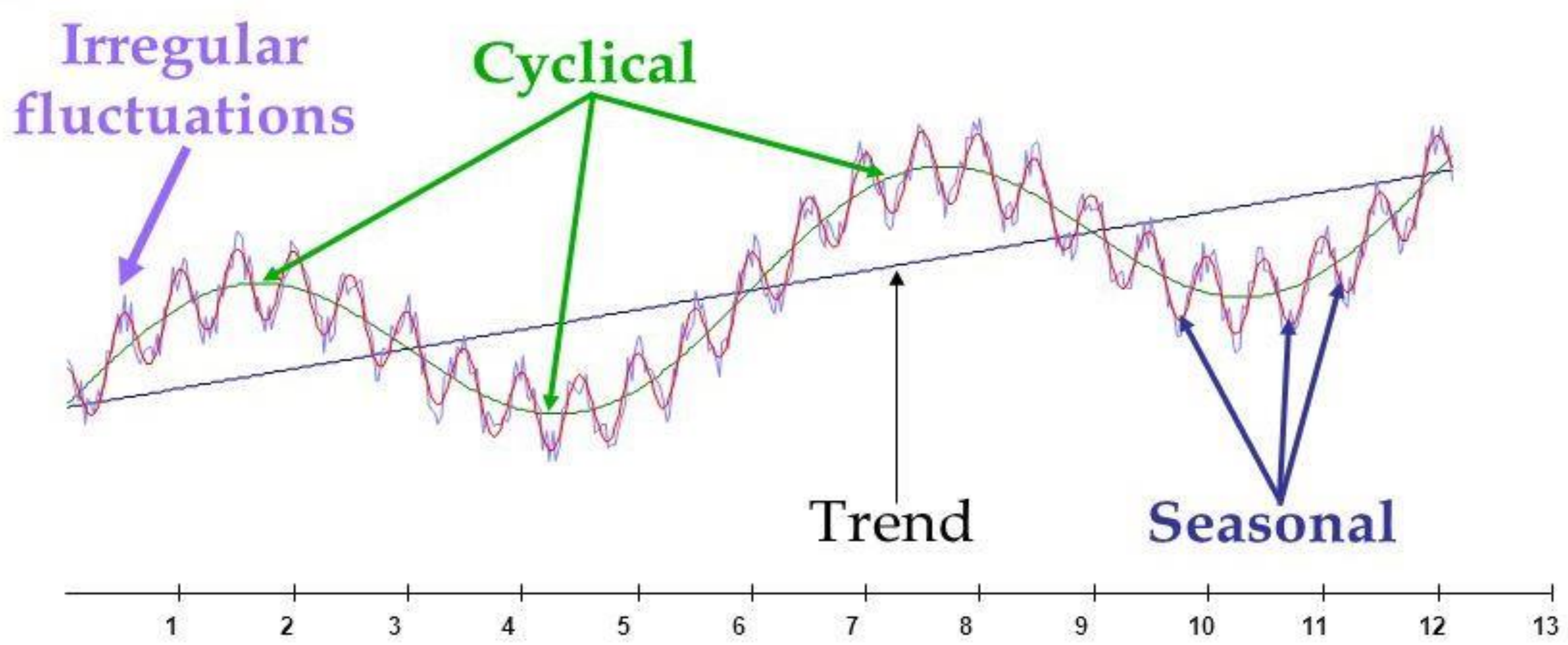
مولفه‌های سری زمانی– غیر معمول

- غیر معمول: تغییراتی که قابل پیش‌بینی نیستند و نظم و الگوی خاصی ندارند.



مولفه‌های سری زمانی– تناوب

- تناوب : تغییرات یکسان و تکراری در مقاطع میان‌مدت، تناوب در سری زمانی نامیده می‌شود.



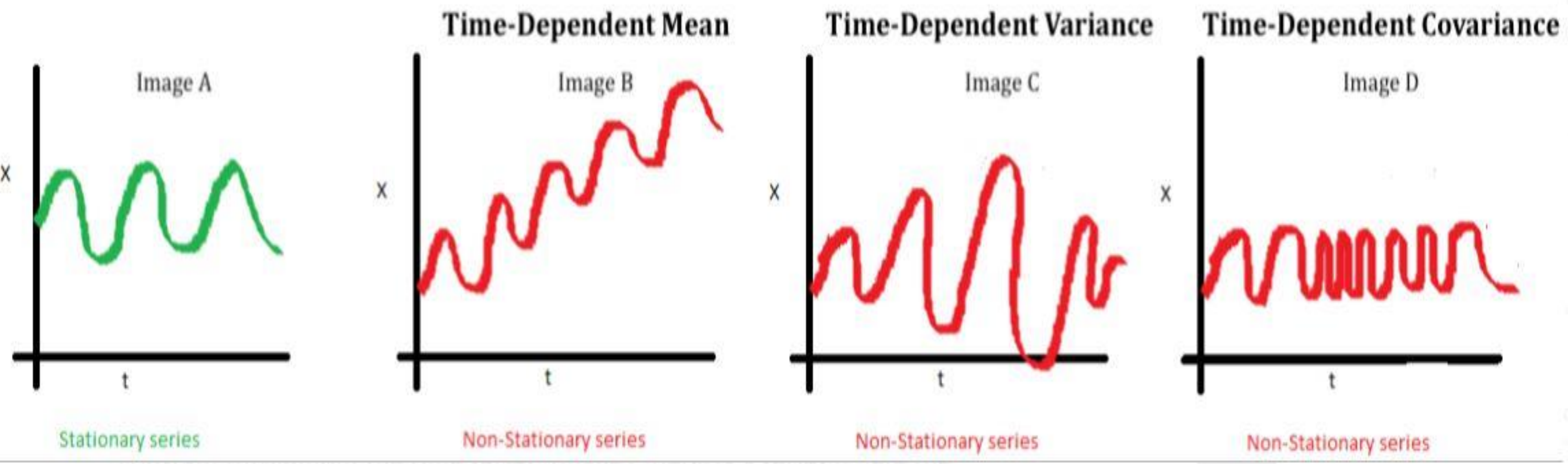
سری زمانی

- سری ایستا: یک سری زمانی است که قوانین حاکم بر تغییرات مقدارها، وابسته به زمان نباشد. قوانین احتمالی حاکم بر فرایند، با زمان تغییر نمی‌کند یا به عبارتی فرایند در تعادل آماری است.
- سری غیر ایستا: سری زمانی است که تغییرات آن در طول زمان متفاوت است و به زمان وابسته است بنابراین اطلاعات آماری ثابتی نخواهد داشت.

پیش‌بینی سری ایستا ساده‌تر است.

سری زمانی ایستا

- میانگین ثابت
- واریانس ثابت : اختلاف بین داده و میانگین وابسته به زمان نباشد.
- کوواریانس مستقل از زمان: تفاضل بین مقادیر ابتدا وانتهای بازه‌های ثابت وابسته به زمان نباشد.



تست ایستایی سری زمانی

- نگاه کردن به نمودار
- نتایج آماری
- تستی به نام **Augmented Dickey-Fuller (ADF)** وجود دارد.
  - تست عددی را به عنوان نتیجه به ما می دهد که اگر صفر یا نزدیک به صفر باشد به معنای ایستا بودن سری است.

مدل های تحلیل سری زمانی

- اگر $X(t)$ مقدار سری زمانی باشد و $T(t)$ اثر مولفه روند، $C(t)$ اثر مولفه تناوب، $S(t)$ اثر مولفه فصل و $I(t)$ نیز اثر مولفه تصادفی باشد، می‌توان این دو مدل سری زمانی را به صورت زیر معرفی کرد:
- **«مدل ضربی» Multiplicative Model:** $X(t)=T(t)\times S(t)\times C(t)\times I(t)$
- **«مدل جمعی» Additive Model :** $X(t)=T(t)+S(t)+C(t)+I(t)$

روش‌های ایستا کردن سری زمانی

- عملگر میانگین متحرک(Moving Average) :
- X' (t) = [X(t) + X(t-1) + … + X(t-(k-1))]/k
- Y(t) = X(t) - X' (t)
- تفاضل (differencing):
- d^1 (t) = X(t) - X(t-1) , d^1 (t-1) = X(t-1) - X(t-2) , …
- d^2 (t) = d^1 (t) - d^1 (t-1) , d^2 (t-1) = d^1 (t-1) - d^1 (t-2) , …

ویژگی خود همبستگی (AutoCorrelation)

- هدف از تحلیل سری زمانی، یافتن رابطه بین داده‌های زمانهای مختلف است.
- خود همبستگی: وابستگی بین مقدار دنباله‌ای برحسب زمان است.
- تابعی که خودهمبستگی را برحسب یک فاصله زمانی بین مشاهدات محاسبه می کند «تابع خودهمبستگی» (Autocorrelation Function) نامیده می‌شود.
- Lag : فاصله بین دو داده در سری زمانی

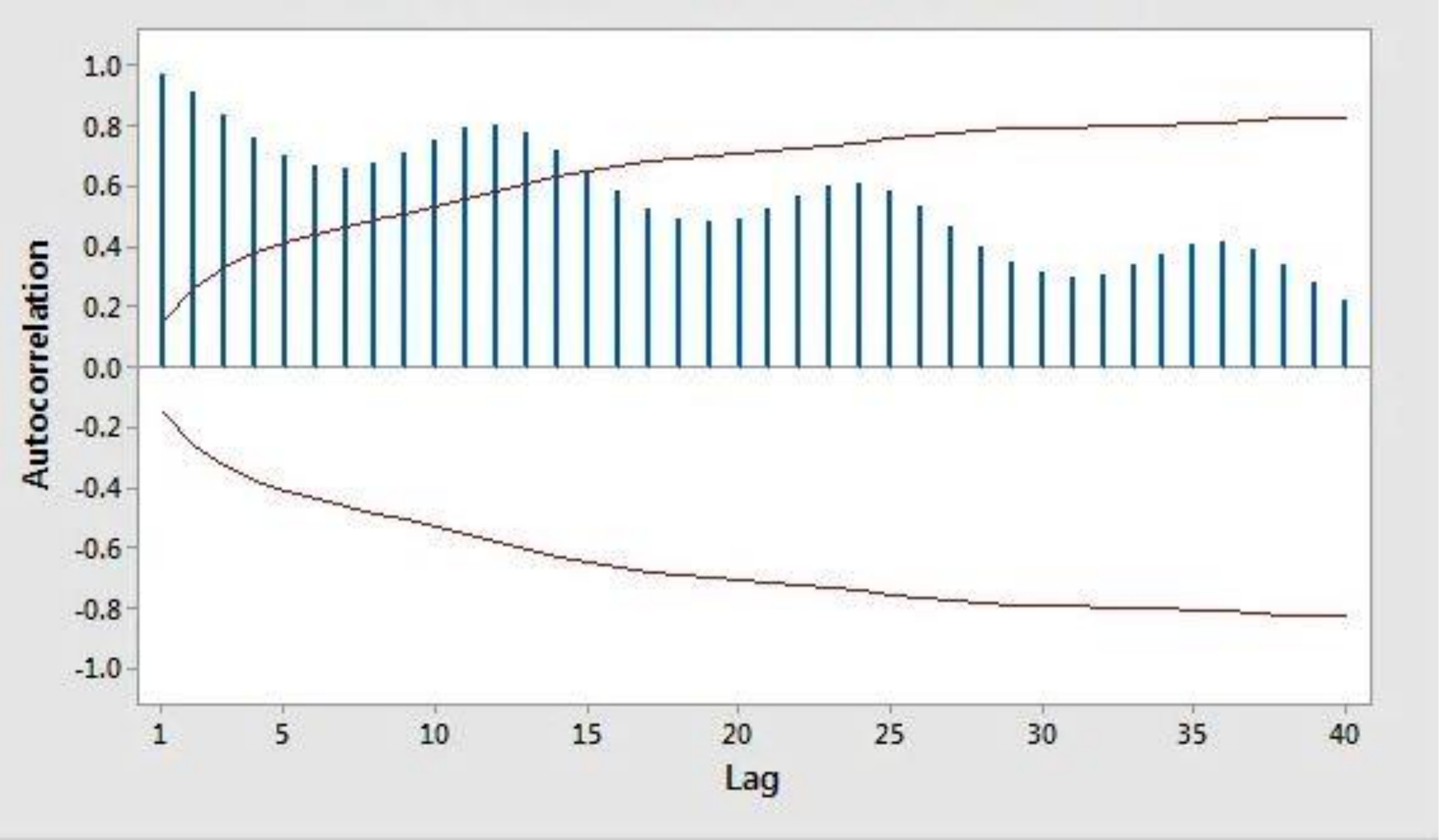
تابع خود همبستگی (AutoCorrelation Function-ACF)

- فرض کنید X_t مقدار سری زمانی در زمان t باشد.
- تابع خودهمبستگی برای این سری زمانی، ضرایب همبستگی بین مشاهدات X_t و X_{t-h} را براساس $h=1,2,3,\cdots$ نشان می‌دهد:

$$ACF = \frac{Cov(X_t, X_{t-h})}{\sigma_t \sigma_t}$$

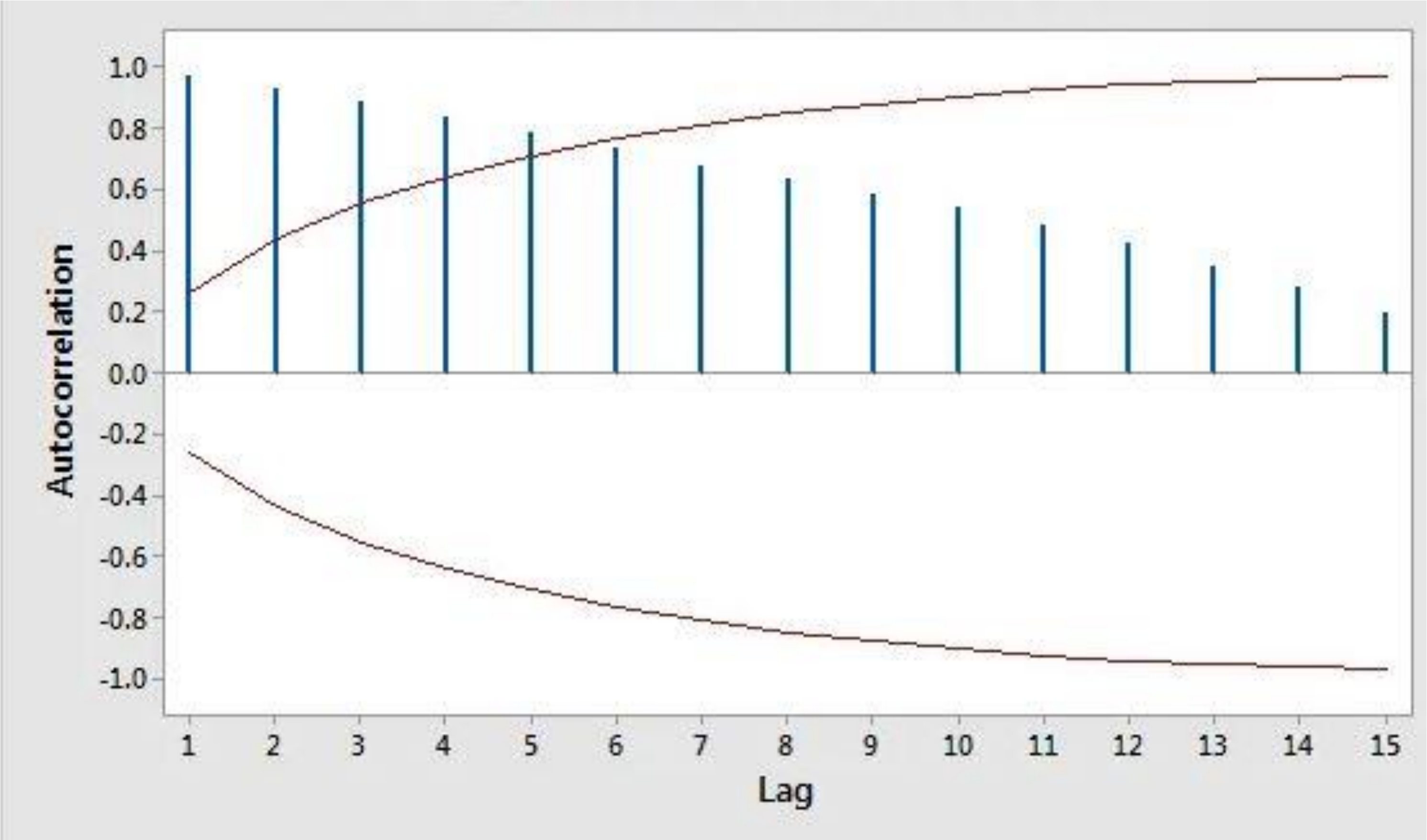
$$ACF = \frac{Cov(X_t, X_{t-h})}{Var(X_t)}$$

خود همبستگی



1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Autocorrelation
1
5
10
15
20
25
30
35
40
Lag

خود همبستگی

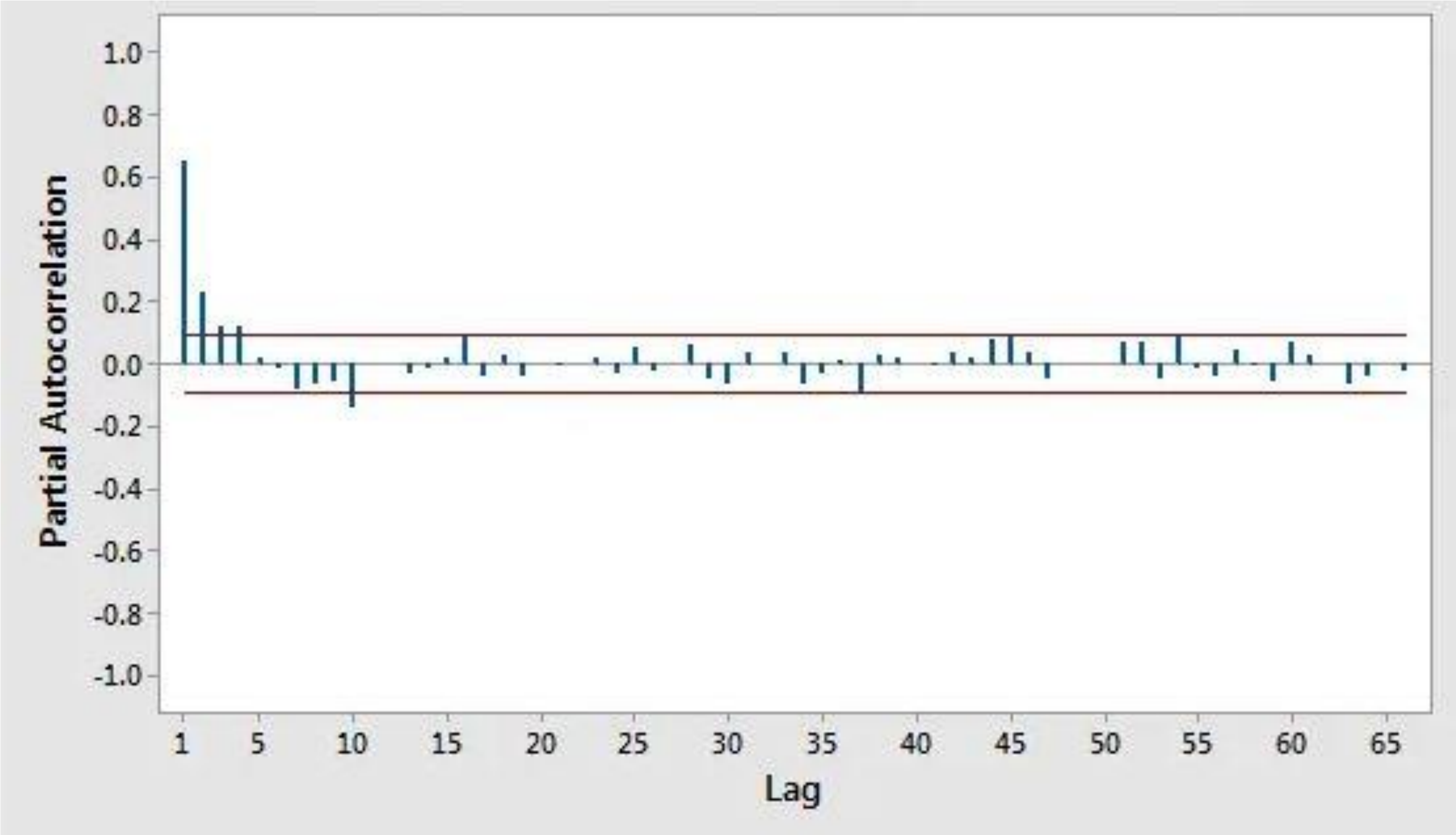


Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Lag

خود همبستگی جزئی (Partial AutoCorrelation)

- وابستگی مستقیم بین دو داده را بیان می کند.
- در خودهمبستگی، وابستگی های غیر مستقیم هم دخیل هستند اما در خودهمبستگی جزئی وابستگی غیر مستقیم حذف می شود.
- این ویژگی رابطه بین دو داده را طوری بیان می کند که با هیچ lag کوچکتری قابل بیان نیست.

خود همبستگی جزئی (Partial AutoCorrelation)



خود همبستگی و خود همبستگی جزئی

- از ویژگی خود همبستگی برای انتخاب مدل سری زمانی استفاده می‌شود.
- از ویژگی خودهمبستگی جزئی برای تعیین پارامترهای مدل استفاده می‌شود.

داده‌ها و تخمین پارامترهای مدل

- داده سری زمانی: X(t)
- تخمین داده‌ها : Y(t)
- مدل : $Y = f(X,t,\alpha)$
- بنابراین برای هر t، دو مقدار X(t) و Y(t) وجود دارد که اولی حاصل اندازه‌گیری از فرآیند تصادفی و دومی از مدل به دست آمده.
- این دو مقدار به دلایلی(مناسب نبودن مدل، نویزو...) با هم تفاوت دارند که به آن مانده (residual) گویند:
- r(t) = Y(t) - X(t)

داده‌ها و تخمین پارامترهای مدل

- در یادگیری ماشین همواره هدف این است که مانده کمینه شود.
- بنابراین تابع هدف شکل می گیرد:
- Min r(t)
- در سری زمانی معمولا از روش کمترین مربعات استفاده می شود:
- $Cost(\alpha) = \sum_t r(t)^2 = \sum_t (X(t) - Y(t))^2 = \sum_t (X(t) - f(X, t, \alpha))^2$
- هدف کمینه کردن تابع Cost هست. پارامترهای α طوری تعیین می‌شوند این تابع کمینه شود.
- می خواهیم نقطه کمینه یک تابع را به دست آوریم؛ از تابع مشتق می گیریم و برابر صفر قرار می دهیم.

داده‌ها و تخمین پارامترهای مدل

- با مشخص شدن پارامترها α ، می توان $Y(t)$ را به دست آورد.
- برای این مرحله که فیت کردن مدل با داده گویند از داده‌های آموزش استفاده می‌شود.
- برای سنجش مدل از داده‌های تست استفاده می شود.
- بسته به کاربرد، روشهای مختلفی برای سنجش استفاده می‌شود.

مدل اتو رگرسیو (AR)

- یک مدل خطی است (؟).
- زمانی قابل استفاده است که داده‌های سری زمانی ویژگی خود همبستگی داشته باشند.
- بر اساس داده‌های زمانی قبلی، داده زمانی فعلی را تخمین می‌زنند.
- مرتبه k: بر اساس k تاخیر (لگ) ، تخمین صورت می گیرد.

مدل اتو رگرسیو (AR)

- مرتبه 1:
- Y(t) = C + $\varphi_1 X(t-1)$
- مرتبه p:
- Y(t) = C + $\varphi_1 X(t-1) + \varphi_2 X(t-2) + \ \ldots + \varphi_p X(t-p) = C + \sum_{i=1}^{p} \varphi_i X(t-i)$
- r(t) = Y(t) - X(t)
- هر چه ضرایب φ_i کوچکتر باشند یعنی وابستگی کمتری بین داده ای که تخمین می زنیم و داده های قبلی وجود دارد و بالعکس.
- $-1 \leq \varphi_i \leq 1$

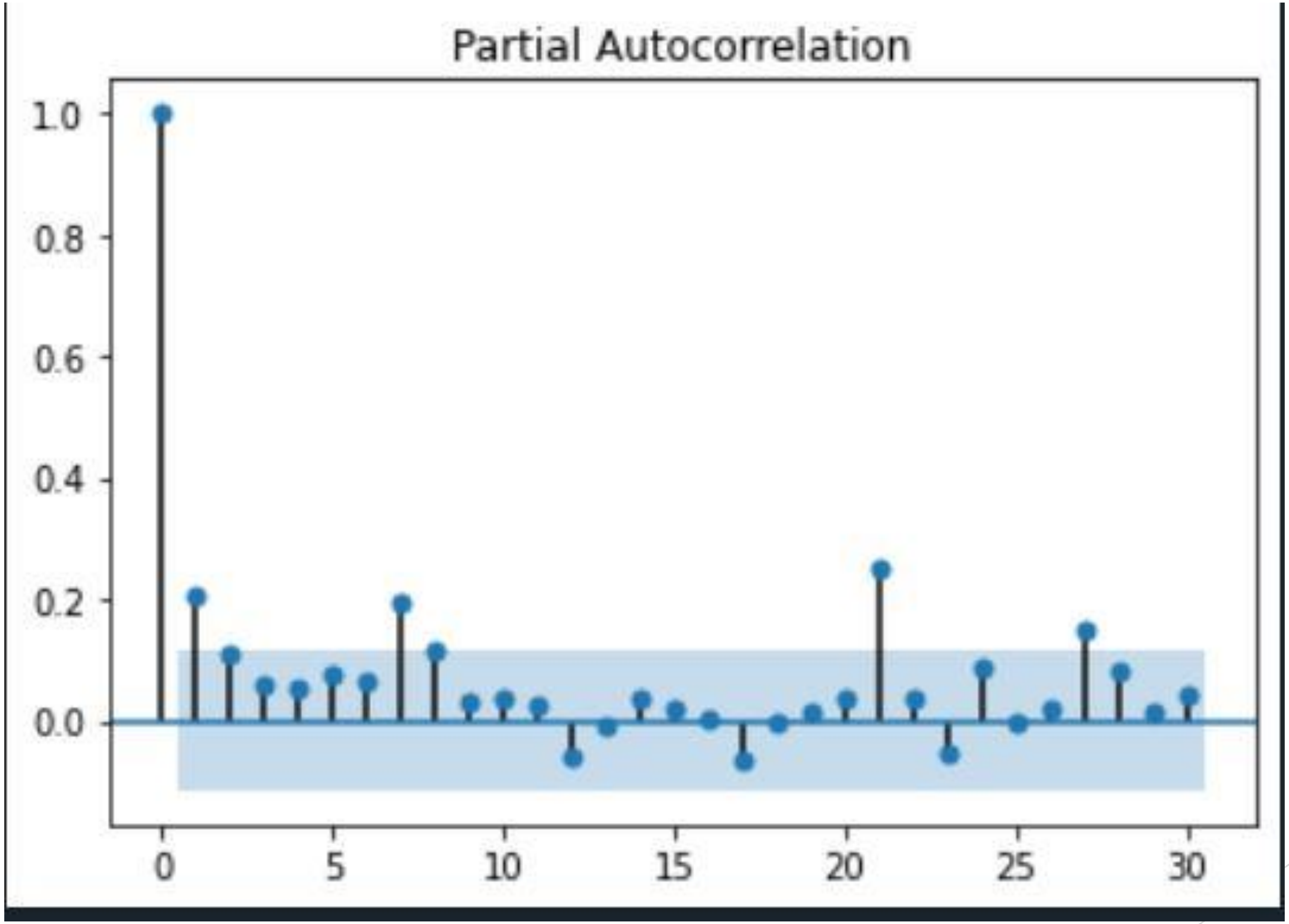
مدل اتو رگرسیو (AR)

- mport pandas as pd
- import matplotlib.pyplot as plt
- from statsmodels.tsa.stattools import adfuller
- from statsmodels.graphics.tsaplots import plot_acf
- from statsmodels.graphics.tsaplots import plot_pacf
- from statsmodels.tsa.ar_model import AutoRegfrom statsmodels.tsa.ar_model import AutoReg

مدل اتو رگرسیو (AR)

- from statsmodels.tsa.ar_model import AutoRegfrom statsmodels.tsa.ar_model import AutoReg
- df = pd.read_csv("BD.csv")
- df['Date'] = pd.to_datetime(df['Date'],)
- df.set_index('Date', inplace = True)
- size = round(len(df)*.8)
- X_tr = df.iloc[:size]
- X_te = df.iloc[size:]
- result = adfuller(X_tr)
- print(result)
- plot_acf(X_tr, lags=30)
- plot_pacf(X_tr, lags=30)
- ar = AutoReg(X_tr, lags =3).fit()
- print (ar.summary())

مدل اتو رگرسیو (AR)



مدل اتو رگرسیو (AR)

result - Tuple (6 elements)

Ind	Type	Size	Value
0	float64	1	-3.41377339551857
1	float64	1	0.010498203476885785
2	int	1	7
3	int	1	284
4	dict	3	{'1%':-3.4535872903895797, '5%':-2.871771355211212, '10%':-2.572221728 ...
5	float64	1	1868.9845862033646

Close

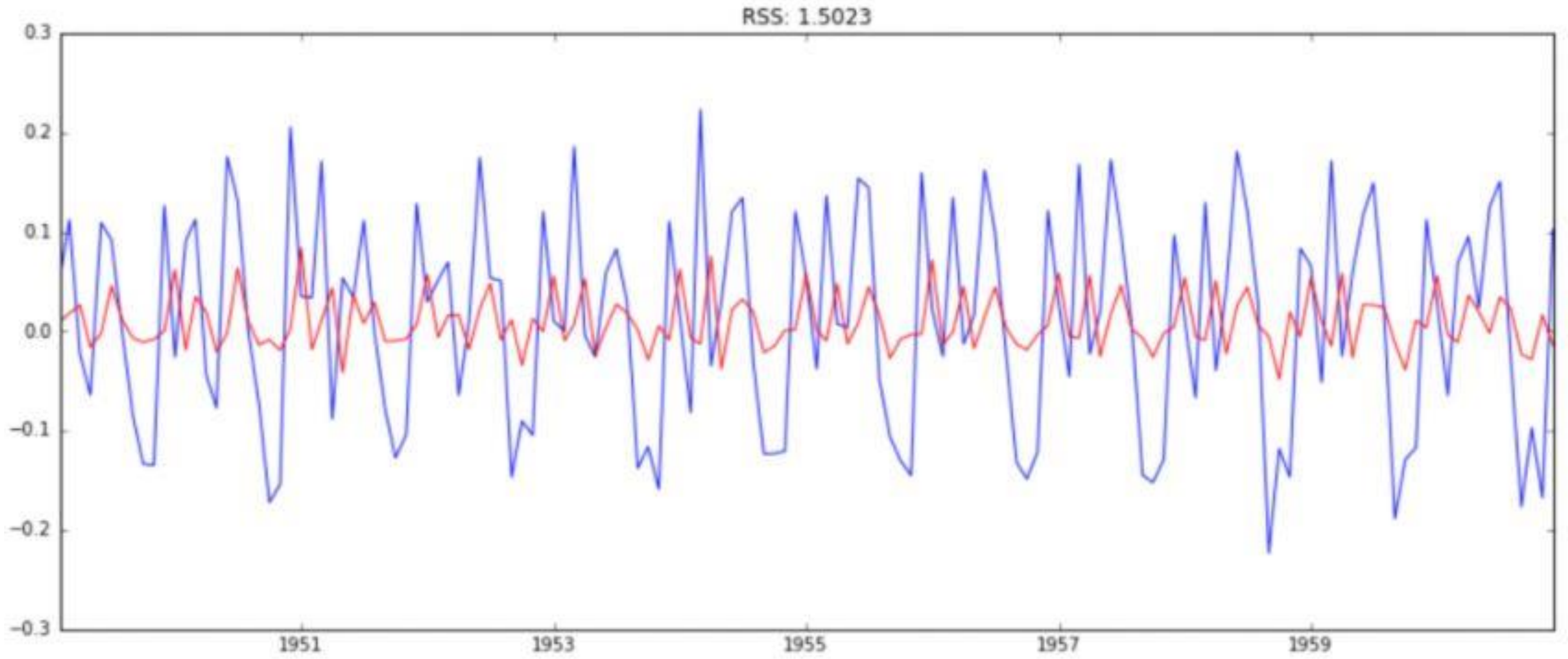
مدل اتو رگرسیو (AR)

```
                            AutoReg Model Results
==============================================================================
Dep. Variable:                 Births   No. Observations:                  292
Model:                     AutoReg(3)   Log Likelihood                -984.610
Method:               Conditional MLE   S.D. of innovations              7.301
Date:                Sun, 13 Mar 2022   AIC                              4.011
Time:                        16:53:25   BIC                              4.074
Sample:                    01-04-1959   HQIC                             4.036
                         - 10-19-1959
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept     28.3384      3.546      7.992      0.000      21.388      35.288
Births.L1      0.1650      0.059      2.817      0.005       0.050       0.280
Births.L2      0.0987      0.059      1.673      0.094      -0.017       0.214
Births.L3      0.0593      0.058      1.015      0.310      -0.055       0.174
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.8339           -0.0000j            1.8339           -0.0000
AR.2           -1.7489           -2.4775j            3.0326           -0.3478
AR.3           -1.7489           +2.4775j            3.0326            0.3478
-----------------------------------------------------------------------------
```

مدل اتو رگرسیو (AR)

```
Dep. Variable:                 Births   No. Observations:                  292
Model:                    AutoReg(7)   Log Likelihood                -962.973
Method:              Conditional MLE   S.D. of innovations               7.099
Date:               Sun, 13 Mar 2022   AIC                               3.983
Time:                       16:59:53   BIC                               4.098
Sample:                   01-08-1959   HQIC                              4.029
                        - 10-19-1959
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept     18.8197      4.599      4.092      0.000       9.805      27.835
Births.L1      0.1453      0.058      2.505      0.012       0.032       0.259
Births.L2      0.0637      0.058      1.093      0.275      -0.051       0.178
Births.L3      0.0307      0.058      0.527      0.598      -0.084       0.145
Births.L4      0.0304      0.058      0.521      0.603      -0.084       0.145
Births.L5      0.0473      0.058      0.813      0.416      -0.067       0.161
Births.L6      0.0397      0.058      0.684      0.494      -0.074       0.154
Births.L7      0.1955      0.058      3.372      0.001       0.082       0.309
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.1371           -0.0000j            1.1371           -0.0000
AR.2            0.7784           -0.9854j            1.2557           -0.1436
AR.3            0.7784           +0.9854j            1.2557            0.1436
AR.4           -1.1650           -0.5672j            1.2958           -0.4279
AR.5           -1.1650           +0.5672j            1.2958            0.4279
AR.6           -0.2835           -1.2723j            1.3035           -0.2849
AR.7           -0.2835           +1.2723j            1.3035            0.2849
-----------------------------------------------------------------------------
```

مدل اتو رگرسیو (AR)



مدل میانگین متحرک(MA)

- در این مدل تاثیر شوکهای تصادفی در نظر گرفته می‌شود.
- فاصله بین مقدارخودهمبستگی و مقدار سری، شوک تصادفی را نشان می‌دهد.
- در نمودار خود همبستگی متوجه می‌شدیم که داده‌ها حدود چند لگ تحت تاثیر یکدیگر هستند.
- از همین ویژگی می توان فهمید که شوکهای ناگهانی تا چند لگ بعدی تاثیر می گذارند.

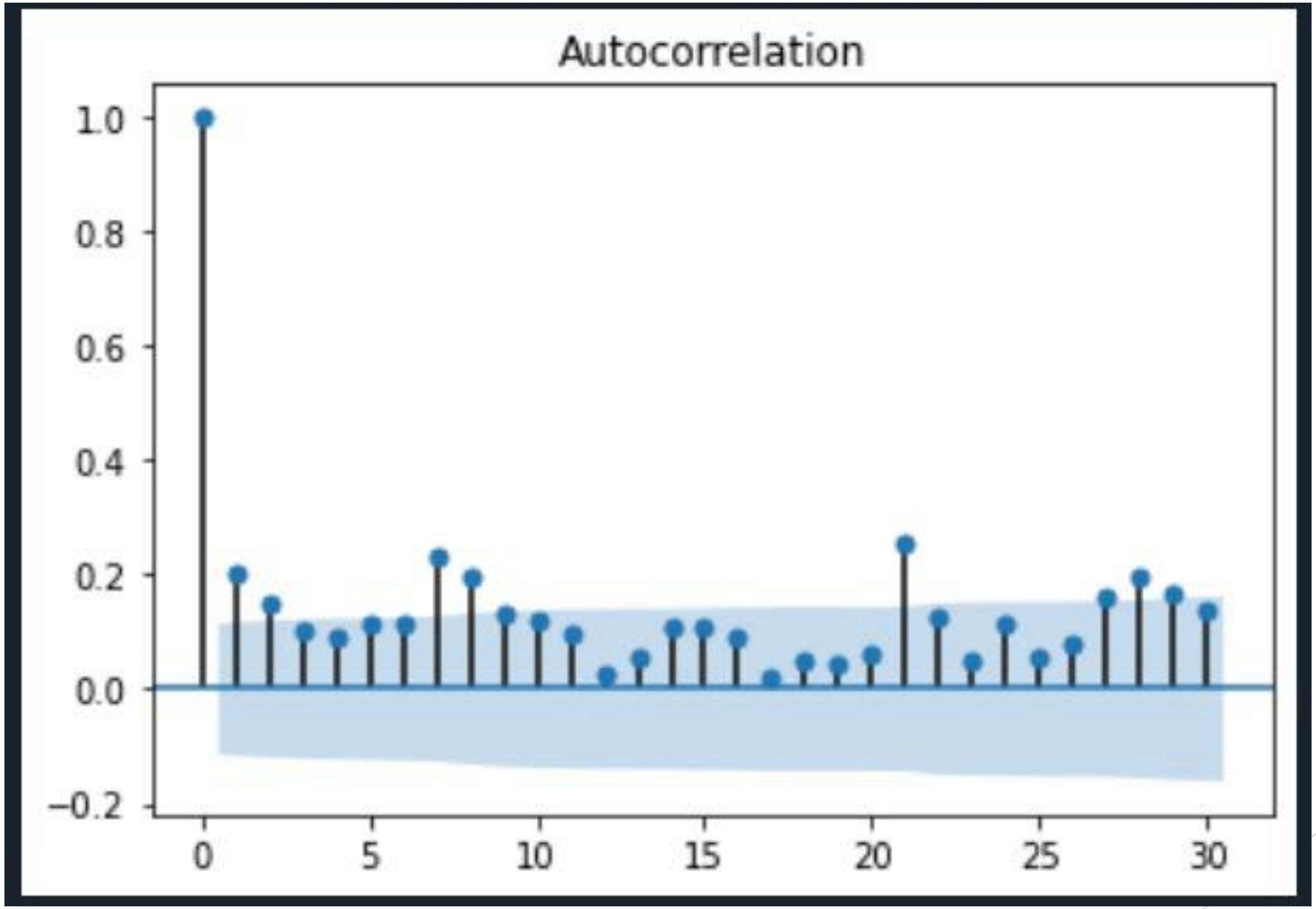
مدل میانگین متحرک(MA)

- مرتبه 1:
- Y(t) = μ + $\theta_1 r(t-1)$
- مرتبه q:
- Y(t) = μ + $\theta_1 r(t-1) + \theta_2 r(t-2) + \ \ldots + \theta_q r(t-q) = \mu + \sum_{i=1}^{q} \theta_i r(t-i)$
- ε_t= Y(t) - X(t)
- هر چه ضرایب θ_i کوچکتر باشند یعنی وابستگی کمتری بین داده ای که تخمین می زنیم و مانده‌های قبلی وجود دارد و بالعکس.
- $-1 \leq \theta_i \leq 1$

مدل میانگین متحرک(MA)

- from statsmodels.tsa.arima_model import ARMA
- ma = ARMA(X_tr, order=(0,3)).fit()
- print(ma.summary())

مدل میانگین متحرک(MA)



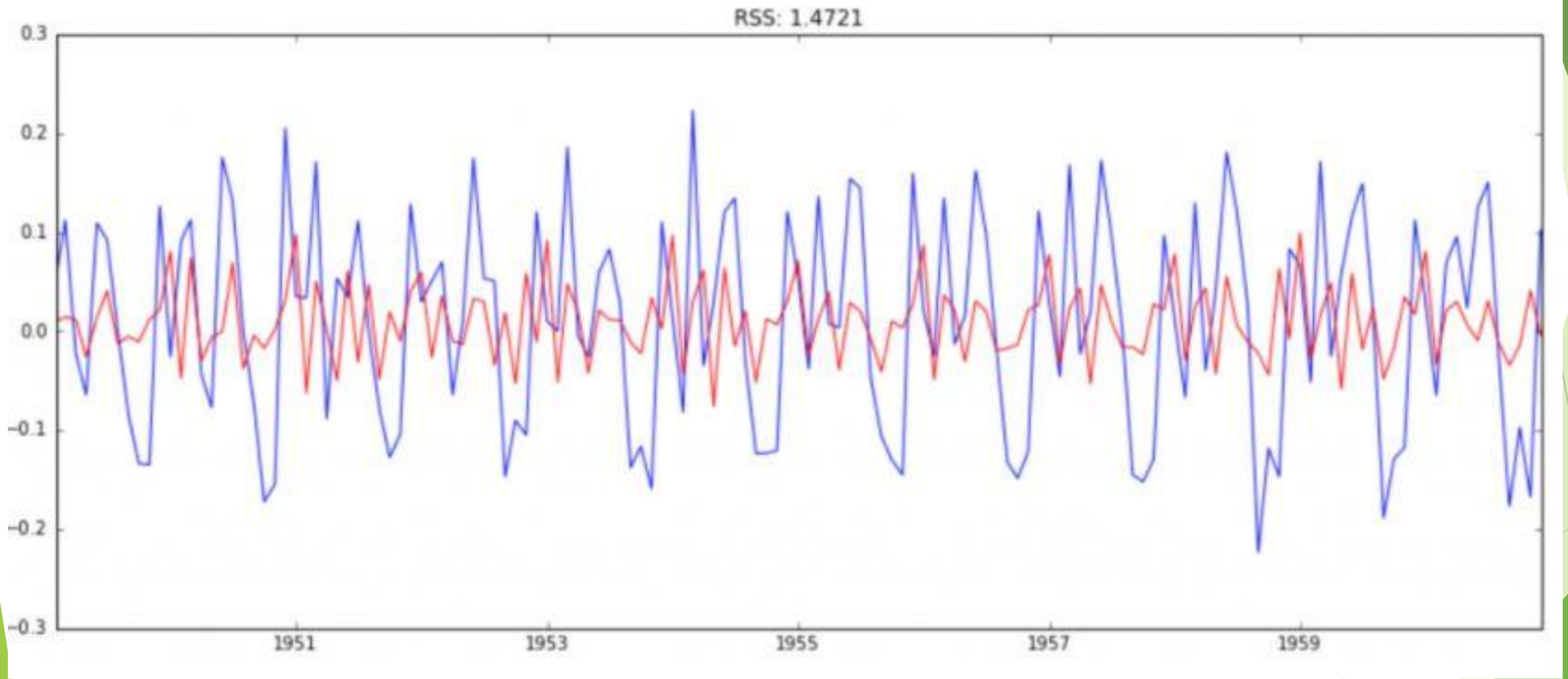
مدل میانگین متحرک(MA)

```
                              ARMA Model Results
==============================================================================
Dep. Variable:                 Births   No. Observations:                  292
Model:                     ARMA(0, 3)   Log Likelihood                -996.002
Method:                       css-mle   S.D. of innovations              7.330
Date:                Sun, 13 Mar 2022   AIC                           2002.005
Time:                        17:16:14   BIC                           2020.388
Sample:                    01-01-1959   HQIC                          2009.368
                         - 10-19-1959
================================================================================
                   coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
const           41.7097      0.584     71.431      0.000      40.565      42.854
ma.L1.Births     0.1724      0.058      2.958      0.003       0.058       0.287
ma.L2.Births     0.1195      0.056      2.123      0.034       0.009       0.230
ma.L3.Births     0.0715      0.060      1.196      0.232      -0.046       0.189
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
MA.1            0.5134           -2.2181j            2.2768           -0.2138
MA.2            0.5134           +2.2181j            2.2768            0.2138
MA.3           -2.6989           -0.0000j            2.6989           -0.5000
-----------------------------------------------------------------------------
```

مدل میانگین متحرک(MA)

```
==============================================================================
Dep. Variable:                 Births   No. Observations:                  292
Model:                     ARMA(0, 7)   Log Likelihood                -991.424
Method:                       css-mle   S.D. of innovations              7.213
Date:                Sun, 13 Mar 2022   AIC                           2000.847
Time:                        17:17:46   BIC                           2033.938
Sample:                    01-01-1959   HQIC                          2014.102
                         - 10-19-1959
================================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
const         41.7029      0.647     64.428      0.000      40.434      42.972
ma.L1.Births   0.1211      0.061      1.995      0.046       0.002       0.240
ma.L2.Births   0.0877      0.062      1.424      0.154      -0.033       0.208
ma.L3.Births   0.0261      0.062      0.418      0.676      -0.096       0.148
ma.L4.Births   0.0248      0.053      0.467      0.640      -0.079       0.129
ma.L5.Births   0.0630      0.054      1.163      0.245      -0.043       0.169
ma.L6.Births   0.0341      0.061      0.563      0.573      -0.085       0.153
ma.L7.Births   0.1844      0.065      2.835      0.005       0.057       0.312
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
MA.1            1.1028           -0.6226j            1.2664           -0.0818
MA.2            1.1028           +0.6226j            1.2664            0.0818
MA.3           -1.2607           -0.0000j            1.2607           -0.5000
MA.4           -0.7846           -1.0023j            1.2728           -0.3557
MA.5           -0.7846           +1.0023j            1.2728            0.3557
MA.6            0.2197           -1.2676j            1.2865           -0.2227
MA.7            0.2197           +1.2676j            1.2865            0.2227
-----------------------------------------------------------------------------
```

مدل میانگین متحرک(MA)



مدل ARMA

- ترکیبی از روش MA(p) و AR(q) هست.
- ARMA(p,q):

- Y(t) = μ + $\sum_{i=1}^{p} \varphi_i X(t-i) + \sum_{i=1}^{q} \theta_i r(t-i)$
- ε_t= Y(t) - X(t)

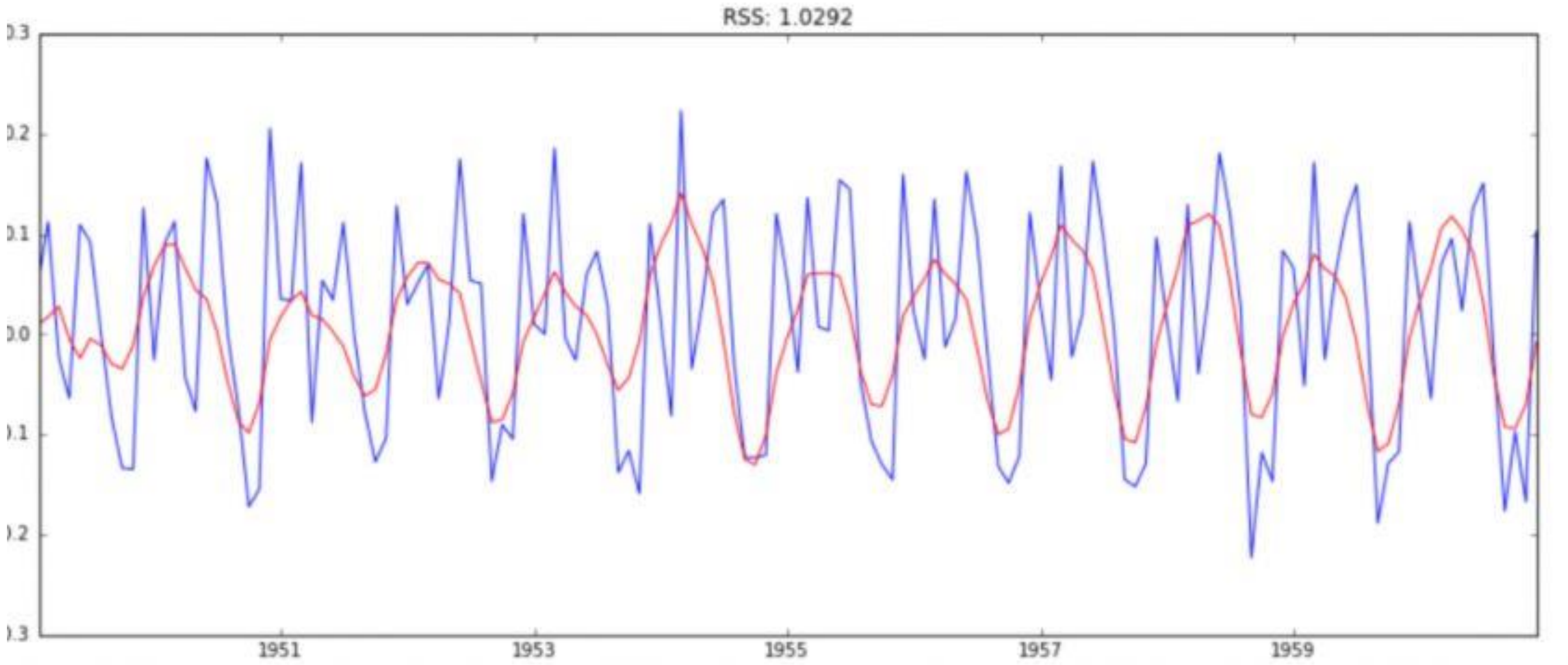
- هر چه ضرایب φ_i کوچکتر باشند یعنی وابستگی کمتری بین داده ای که تخمین می زنیم و داده های قبلی وجود دارد و بالعکس.
- $-1 \leq \varphi_i \leq 1$

مدل ARMA

- ترکیبی از روش MA(p) و AR(q) هست.
- ARMA(p,q):

- Y(t) = $C + \sum_{i=1}^{p} \varphi_i X(t-i) + \sum_{i=1}^{q} \theta_i r(t-i)$
- $ARMA(2,1) = C + \varphi_1 X(t-1) + \varphi_2 X(t-2) + \theta_1 r(t-1)$

مدل ARMA



مدل ARIMA


Differentiation
Actual Series
Series After Differentiation

مدل ARIMA

- ترکیبی از روش MA(p) و AR(q) و Dif(d) هست.
- ARIMA(p,d,q):
- $ARIMA(2,0,1) = C + \varphi_1 X(t-1) + \varphi_2 X(t-2) + \theta_1 r(t-1)$