



# Giới thiệu mô hình ARIMA

Bởi:

Phạm Trí Cao

## Tính dừng của dữ liệu

Quá trình ngẫu nhiên(Stochastic process)

Bất cứ dữ liệu chuỗi thời gian nào cũng được tạo ra bằng một quá trình ngẫu nhiên. Một dãy số liệu thực tế cụ thể như giá bắp cải từng tháng ở hình 7.1 là kết quả của một quá trình ngẫu nhiên. Đối với dữ liệu chuỗi thời gian, chúng ta có những khái niệm về tổng thể và mẫu như sau:

Quá trình ngẫu nhiên là một tổng thể.

Số liệu thực tế sinh ra từ quá trình ngẫu nhiên là mẫu.

Tính dừng(Stationary)

Một quá trình ngẫu nhiên được gọi là có tính dừng khi nó có các tính chất sau:

Kỳ vọng không đổi theo thời gian,  $E(Y_t) = \mu$ .

Phương sai không đổi theo thời gian,  $\text{Var}(Y_t) = E(Y_t - \mu) = \sigma^2$ .

Đồng phương sai chỉ phụ thuộc khoảng cách của độ trễ mà không phụ thuộc thời điểm tính đồng phương sai đó,  $v_k = E[(Y_t - \mu)(Y_{t-k} - \mu)]$  không phụ thuộc  $t$ .

Lưu ý: Chúng ta có thể biến dữ liệu chuỗi thời gian từ không có tính dừng thành có tính dừng bằng cách lấy sai phân của nó.

$w_t = Y_t - Y_{t-1}$ : Sai phân bậc nhất

$w_t^2 = w_t - w_{t-1}$ : Sai phân bậc hai...

## Hàm tự tương quan và hàm tự tương quan mẫu

Hàm tự tương quan(ACF) ở độ trễ  $k$  được ký hiệu là  $\rho_k$  được định nghĩa như sau:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\mathbb{E}[(Y_t - \mu)(Y_{t-k} - \mu)]}{\mathbb{E}[(Y_t - \mu)^2]}$$

(7.11)

Tính chất của ACF

$\rho_k$  không có thứ nguyên.

Giá trị của  $\rho_k$  nằm giữa -1 và 1.

Trong thực tế chúng ta chỉ có thể có số liệu thực tế là kết quả của quá trình ngẫu nhiên, do đó chúng chỉ có thể tính toán được hàm tự tương quan mẫu(SAC), ký hiệu là  $r_k$ .

$$r_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$$

với

$$\hat{\gamma}_k = \frac{\sum (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{n}$$

và

$$\hat{\gamma}_0 = \frac{\sum (Y_t - \bar{Y})^2}{n}$$

Độ lệch chuẩn hệ số tự tương quan mẫu

$s(r_j) =$

$$\frac{\sqrt{1 + 2 \sum_{i=1}^H r_i^2}}{\sqrt{n}}$$

(7.12)

Trị thống kê t

$$t_k = \frac{r_k}{s(r_k)} \quad (7.13)$$

Với cỡ mẫu lớn thì  $t_k \sim Z$  nên với  $t > 1,96$  thì  $r_k$  khác không có ý nghĩa thống kê, khi đó người ta gọi  $r_k$  là 1 đỉnh.

## Giới thiệu mô hình ARIMA

Các phần mềm kinh tế lượng sẽ tính toán cho chúng ta kết quả của SAC và các giá trị đến hạn (hoặc trị thống kê t) của nó ứng với mức ý nghĩa  $\alpha = 5\%$ .

Thống kê Ljung-Box

$$LB = n(n+2) \sum_{k=1}^m \left( \frac{r_k^2}{n-k} \right) \sim \chi_m^2$$

(7.14)

n là cỡ mẫu

m là chiều dài của độ trễ

$H_0$ : Tất cả các  $r_k$  đều bằng 0.

$H_1$ : Không phải tất cả các  $r_k$  đều bằng 0.

Nếu  $LB > \chi_{m,1-\alpha}^2$  thì ta bác bỏ  $H_0$ .

Một số phần mềm kinh tế lượng có tính toán trị thống kê LB.

## Hàm tự tương quan riêng phần (PACF)

Hệ số tự tương quan riêng phần với độ trễ k đo lường tương quan của  $Y_{t-k}$  với  $Y_t$  sau khi loại trừ tác động tương quan của tất cả các độ trễ trung gian. Công thức tính PACF như sau

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-j} r_j}{1 - \sum_{j=1}^{k-1} r_{k-j} r_j}$$

(7.15)

Độ lệch chuẩn của  $r_{kk}$

Công thức tính độ lệch chuẩn của  $r_{kk}$  phụ thuộc vào bậc của sai phân. Công thức trình bày ở trên là công thức gần đúng với số quan sát đủ lớn.

$$s(r_{kk}) = \frac{1}{\sqrt{n}} \quad (7.16)$$

## Giới thiệu mô hình ARIMA

Trị thống kê t

$$t_{kk} = \frac{r_{kk}}{s(r_{kk})} \quad (7.17)$$

Với cỡ mẫu lớn thì  $t_{kk} \sim Z$  nên với  $t_{kk} > 1,96$  thì  $r_{kk}$  khác không có ý nghĩa thống kê, khi đó người ta gọi  $r_{kk}$  là 1 đỉnh.

Các chương trình kinh tế lượng có thể tính toán cho chúng ta các giá trị PACF, các giá trị tới hạn hay trị thống kê t.

Mô hình AR, MA và ARMA

Xét quá trình ngẫu nhiên có tính dừng với dữ liệu chuỗi thời gian  $Y_t$  có  $E(Y_t) = \mu$  và sai số ngẫu nhiên  $\mu_t$  có trung bình bằng 0 và phương sai  $\sigma^2$  (nhiều trắng).

Mô hình tự hồi quy (AR-Autoregressive Model)

Mô hình tự hồi quy bậc p được ký hiệu là AR(p) có dạng

$$\begin{aligned} (Y_t - \mu) &= \alpha_1(Y_{t-1} - \mu) + \alpha_2(Y_{t-2} - \mu) + \dots + \alpha_p(Y_{t-p} - \mu) + \varepsilon_t \\ Y_t &= \mu(1 - \alpha_1 - \alpha_2 - \dots - \alpha_p) + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t \quad (7.17) \end{aligned}$$

*Nhận dạng mô hình AR(p):* PACF có đỉnh đến độ trễ p và SAC suy giảm nhanh ngay sau độ trễ thứ nhất thì mô hình dự báo có dạng tự hồi quy bậc p.

Mô hình trung bình trượt(MA-Moving average Model)

Mô hình trung bình trượt bậc q được ký hiệu là MA(q) có dạng

$$Y_t = \mu + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} \quad (7.18)$$

với  $\mu$  là hằng số,  $\varepsilon_t$  là nhiễu trắng.

*Nhận dạng mô hình MA(q):* SAC có đỉnh đến độ trễ q và SPAC suy giảm nhanh ngay sau độ trễ thứ nhất.

Mô hình kết hợp tự hồi quy kết hợp trung bình trượt(ARMA)

## Giới thiệu mô hình ARIMA

Mô hình có tự hồi quy bậc  $p$  và trung bình trượt bậc  $q$  được ký hiệu là ARMA( $p,q$ ) có dạng

$$Y_t = \delta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \dots + \beta_q \epsilon_{t-q}$$

(7.19)

Nhận dạng mô hình ARMA( $p,q$ ): cả SAC và SPAC đều có giá trị giảm dần theo hàm mũ. Nhận dạng đúng  $p$  và  $q$  đòi hỏi phải có nhiều kinh nghiệm. Trong thực hành người ta chọn một vài mô hình ARMA và lựa chọn mô hình tốt nhất.

## Mô hình ARIMA và SARIMA

### ARIMA

Đa số dữ liệu kinh tế theo chuỗi thời gian không có tính dừng(stationary) mà có tính kết hợp(integrated). Để nhận được dữ liệu có tính dừng, chúng ta phải sử dụng sai phân của dữ liệu.

#### Các bậc sai phân

Sai phân bậc 0 là  $I(0)$ : chính là dữ liệu gốc  $Y_t$ .

Sai phân bậc 1 là  $I(1)$ :  $w_t = Y_t - Y_{t-1}$ .

Sai phân bậc 2 là  $I(2)$ :  $w_t^2 = w_t - w_{t-1} \dots$

Sai phân bậc  $d$  ký hiệu  $I(d)$ .

Mô hình ARMA( $p,q$ ) áp dụng cho  $I(d)$  được gọi là mô hình ARIMA( $p,d,q$ ).

### SARIMA

Trong mô hình ARIMA nếu chúng ta tính toán sai phân bậc nhất với độ trễ lớn hơn 1 để khử tính mùa vụ như sau  $w_t = Y_t - Y_{t-s}$ , với  $s$  là số kỳ giữa các mùa thì mô hình được gọi là SARIMA hay ARIMA có tính mùa vụ.

#### Phương pháp luận Box-Jenkins

Phương pháp luận Box-Jenkins cho mô hình ARIMA có bốn bước như sau:

Bước 1: Xác lập mô hình ARIMA( $p,d,q$ )

## Giới thiệu mô hình ARIMA

Dùng các đồ thị để xác định bậc sai phân cần thiết để đồ thị có tính dừng. Giả sử dữ liệu dừng ở  $I(d)$ . Dùng đồ thị SAC và SPAC của  $I(d)$  để xác định  $p$  và  $q$ .

Triển khai dạng của mô hình.

Bước 2: Tính toán các tham số của mô hình.

Trong một số dạng ARIMA đơn giản chúng ta có thể dùng phương pháp bình phương tối thiểu. Một số dạng ARIMA phức tạp đòi hỏi phải sử dụng các ước lượng phi tuyến. Chúng ta không phải lo lắng về việc ước lượng tham số vì các phần mềm kinh tế lượng sẽ tính giúp chúng ta. Quay lại bước 1 xây dựng mô hình với cặp  $(p,q)$  khác đường như cũng phù hợp. Giả sử chúng ta ước lượng được  $m$  mô hình ARIMA.

Bước 3: Kiểm tra chẩn đoán

So sánh các mô hình ARIMA đã ước lượng với các mô hình truyền thống (tuyến tính, đường xu hướng, san bằng số mũ,...) và giữa các mô hình ARIMA với nhau để chọn mô hình tốt nhất.

Bước 4: Dự báo

Trong đa số trường hợp mô hình ARIMA cho kết quả dự báo ngắn hạn đáng tin cậy nhất trong các phương pháp dự báo. Tuy nhiên giới hạn của của ARIMA là:

Số quan sát cần cho dự báo phải lớn.

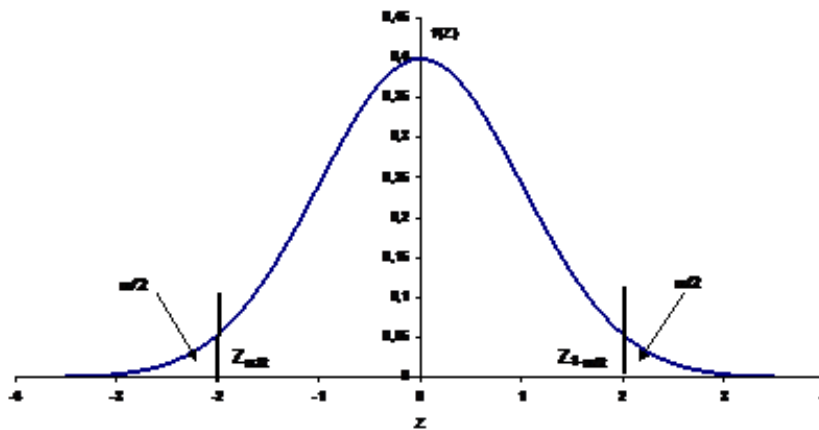
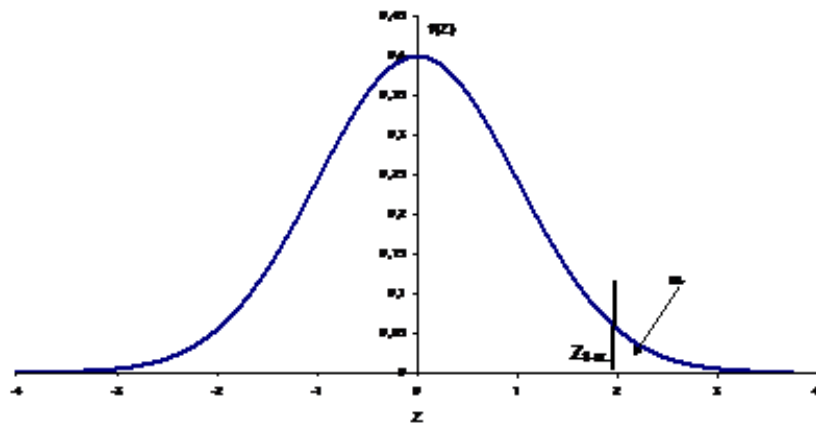
Chỉ dùng để dự báo ngắn hạn

Không thể đưa các yếu tố thay đổi có ảnh hưởng đến biến số cần dự báo của thời kỳ cần dự báo vào mô hình.

Xây dựng mô hình ARIMA theo phương pháp luận Box-Jenkins có tính chất nghệ thuật hơn là khoa học, hơn nữa kỹ thuật và khối lượng tính toán khá lớn nên đòi hỏi phải có phần mềm kinh tế lượng chuyên dùng.

**MỘT SỐ GIÁ TRỊ Z THƯỜNG ĐƯỢC SỬ DỤNG**

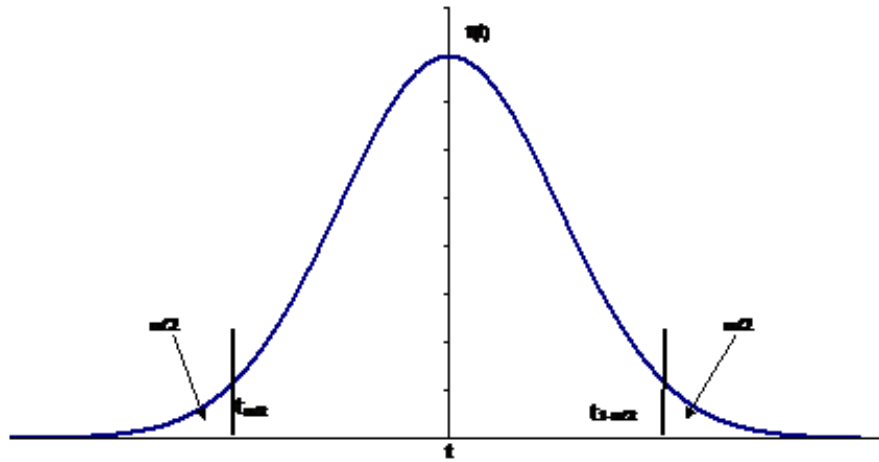
Giới thiệu mô hình ARIMA



Mức ý nghĩa	Kiểm định 1 đuôi	Kiểm định 2 đuôi
$\alpha$	$Z_{1-\alpha}$	$Z_{1-\alpha/2}$
1%	2,326	2,576
5%	1,645	1,960
10%	1,282	1,645
20%	0,842	1,282

Nguồn: hàm Normsinv của Excel.

MỘT SỐ GIÁ TRỊ t THƯỜNG ĐƯỢC SỬ DỤNG

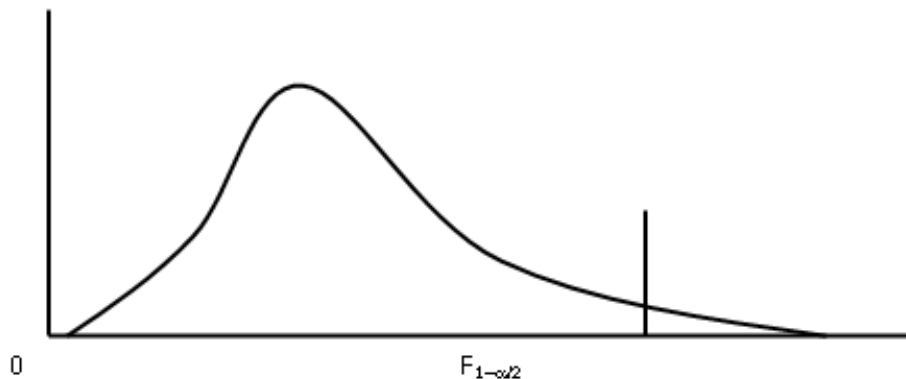


Bậc tự do	Mức ý nghĩa $\alpha$			
	1%	5%	10%	20%
1	63,656	12,706	6,314	3,078
2	9,925	4,303	2,920	1,886
3	5,841	3,182	2,353	1,638
4	4,604	2,776	2,132	1,533
5	4,032	2,571	2,015	1,476
6	3,707	2,447	1,943	1,440
7	3,499	2,365	1,895	1,415

Nguồn: hàm Tinv của Excel.

MỘT SỐ GIÁ TRỊ F TỐI HẠN TRÊN THƯỜNG ĐƯỢC SỬ DỤNG

Mức ý nghĩa  $\alpha = 5\%$





Giới thiệu mô hình ARIMA

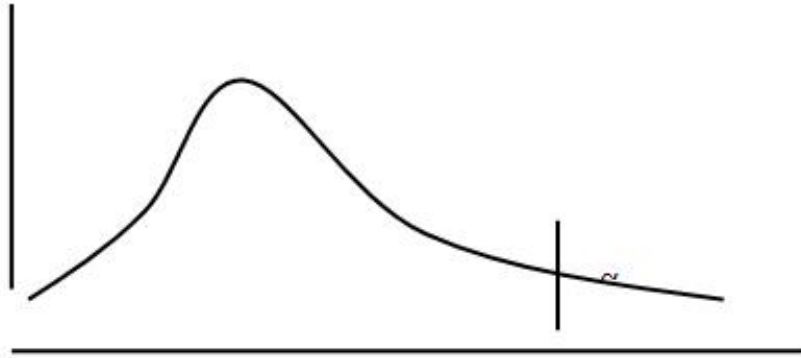
	df1									
df2	1	2	3	4	5	6	7	8	9	10
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
31	4,16	3,30	2,91	2,68	2,52	2,41	2,32	2,25	2,20	2,15
32	4,15	3,29	2,90	2,67	2,51	2,40	2,31	2,24	2,19	2,14
33	4,14	3,28	2,89	2,66	2,50	2,39	2,30	2,23	2,18	2,13
34	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23	2,17	2,12
35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11
36	4,11	3,26	2,87	2,63	2,48	2,36	2,28	2,21	2,15	2,11
37	4,11	3,25	2,86	2,63	2,47	2,36	2,27	2,20	2,14	2,10
38	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09
39	4,09	3,24	2,85	2,61	2,46	2,34	2,26	2,19	2,13	2,08
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08

Giới thiệu mô hình ARIMA

Nguồn: hàm Finv của Excel.

MỘT SỐ GIÁ TRỊ TỚI HẠN TRÊN THƯỜNG ĐƯỢC SỬ DỤNG

Mức ý nghĩa  $\alpha = 5\%$



Giới thiệu mô hình ARIMA

9	21,87	18,92	14,68	12,24
1				
0	23,21	18,31	15,98	13,44
1				
1	24,73	19,68	17,28	14,68
1				
2	26,22	21,03	18,55	15,81
1				
3	27,69	22,36	19,81	16,95
1				
4	29,14	23,68	21,06	18,15
1				
5	30,58	25,00	22,31	19,31
1				
6	32,00	26,30	23,54	20,47
1				
7	33,41	27,59	24,77	21,61
1				
8	34,81	28,87	25,98	22,76
1				
9	36,19	30,14	27,20	23,90
2				
0	37,57	31,41	28,41	25,04
2				
1	38,93	32,67	29,62	26,17
2				
2	40,29	33,92	30,81	27,30
2				
3	41,64	35,17	32,01	28,43
2				
4	42,98	36,42	33,20	29,55
2				
5	44,31	37,65	34,38	30,68
2				
6	45,64	38,88	35,56	31,79
2				
7	46,96	40,11	36,74	32,91
2				
8	48,28	41,34	37,92	34,03
2				
9	49,59	42,56	39,09	35,14
2				
0	50,89	43,77	40,26	36,25
2				
1	52,19	44,98	41,42	37,36
2				
2	53,49	46,18	42,58	38,47
2				
3	54,78	47,40	43,75	39,57
2				
4	56,06	48,60	44,90	40,68
2				
5	57,34	49,80	46,06	41,79
2				
6	58,62	51,00	47,21	42,89
2				
7	59,89	52,19	48,36	43,99
2				
8	61,16	53,38	49,51	45,09
2				
9	62,43	54,57	50,66	46,17
4				
0	63,69	55,76	51,81	47,27

Giới thiệu mô hình ARIMA

Nguồn: Hàm Chiinv của Excel