

Example: MA(2) Model: $y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}$

1. Autocovariance Function of MA(2) Process:

$$\gamma(\tau) = \begin{cases} (1 + \theta_1^2 + \theta_2^2)\sigma_\epsilon^2, & \text{for } \tau = 0, \\ (\theta_1 + \theta_1\theta_2)\sigma_\epsilon^2, & \text{for } \tau = 1, \\ \theta_2\sigma_\epsilon^2, & \text{for } \tau = 2, \\ 0, & \text{otherwise.} \end{cases}$$

2. let $-1/\beta_1$ and $-1/\beta_2$ be two solutions of x from $\theta(x) = 0$.

For invertibility condition, both β_1 and β_2 should be less than one in absolute value.

Then, the MA(2) model is represented as:

$$y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}$$

$$= (1 + \theta_1 L + \theta_2 L^2) \epsilon_t$$

$$= (1 + \beta_1 L)(1 + \beta_2 L) \epsilon_t$$

AR(∞) representation of the MA(2) model is given by:

$$\begin{aligned} \epsilon_t &= \frac{1}{(1 + \beta_1 L)(1 + \beta_2 L)} y_t \\ &= \left(\frac{\beta_1 / (\beta_1 - \beta_2)}{1 + \beta_1 L} + \frac{-\beta_2 / (\beta_1 - \beta_2)}{1 + \beta_2 L} \right) y_t \end{aligned}$$

3. Likelihood Function:

$$f(y_1, y_2, \dots, y_T) = \frac{1}{(2\pi)^{T/2}} |V|^{-1/2} \exp\left(-\frac{1}{2} Y' V^{-1} Y\right)$$

where

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}, \quad V = \sigma_\epsilon^2 \begin{pmatrix} 1 + \theta_1^2 + \theta_2^2 & \theta_1 + \theta_1\theta_2 & \theta_2 & & & 0 \\ \theta_1 + \theta_1\theta_2 & 1 + \theta_1^2 + \theta_2^2 & \theta_1 + \theta_1\theta_2 & \ddots & & \\ \theta_2 & \theta_1 + \theta_1\theta_2 & \ddots & \ddots & & \theta_2 \\ & \ddots & \ddots & 1 + \theta_1^2 + \theta_2^2 & \theta_1 + \theta_1\theta_2 & \\ 0 & & \theta_2 & \theta_1 + \theta_1\theta_2 & 1 + \theta_1^2 + \theta_2^2 & \end{pmatrix}$$

4. **MA(2) +drift:** $y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}$

Mean:

$$y_t = \mu + \theta(L)\epsilon_t,$$

where $\theta(L) = 1 + \theta_1L + \theta_2L^2$.

Therefore,

$$E(y_t) = \mu + \theta(L)E(\epsilon_t) = \mu$$

Example: MA(q) Model: $y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q}$

1. Mean of MA(q) Process:

$$E(y_t) = E(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q}) = 0$$

2. Autocovariance Function of MA(q) Process:

$$\gamma(\tau) = \begin{cases} \sigma_\epsilon^2(\theta_0\theta_\tau + \theta_1\theta_{\tau+1} + \cdots + \theta_{q-\tau}\theta_q) = \sigma_\epsilon^2 \sum_{i=0}^{q-\tau} \theta_i\theta_{\tau+i}, & \tau = 1, 2, \dots, q, \\ 0, & \tau = q + 1, q + 2, \dots, \end{cases}$$

where $\theta_0 = 1$.

3. MA(q) process is stationary.

4. **MA(q) +drift:** $y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q}$

Mean:

$$y_t = \mu + \theta(L)\epsilon_t,$$

where $\theta(L) = 1 + \theta_1L + \theta_2L^2 + \cdots + \theta_qL^q$.

Therefore, we have:

$$E(y_t) = \mu + \theta(L)E(\epsilon_t) = \mu.$$

7.5 ARMA Model

ARMA (Autoregressive Moving Average, 自己回帰移動平均) Process

1. ARMA(p, q)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q},$$

which is rewritten as:

$$\phi(L)y_t = \theta(L)\epsilon_t,$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$.

2. Likelihood Function:

The variance-covariance matrix of Y , denoted by V , has to be computed.

Example: ARMA(1,1) Process: $y_t = \phi_1 y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}$

Obtain the autocorrelation coefficient.

The mean of y_t is to take the expectation on both sides.

$$E(y_t) = \phi_1 E(y_{t-1}) + E(\epsilon_t) + \theta_1 E(\epsilon_{t-1}),$$

where the second and third terms are zeros.

Therefore, we obtain:

$$E(y_t) = 0.$$

The autocovariance of y_t is to take the expectation, multiplying $y_{t-\tau}$ on both sides.

$$E(y_t y_{t-\tau}) = \phi_1 E(y_{t-1} y_{t-\tau}) + E(\epsilon_t y_{t-\tau}) + \theta_1 E(\epsilon_{t-1} y_{t-\tau}).$$

Each term is given by:

$$E(y_t y_{t-\tau}) = \gamma(\tau), \quad E(y_{t-1} y_{t-\tau}) = \gamma(\tau - 1),$$

$$E(\epsilon_t y_{t-\tau}) = \begin{cases} \sigma_\epsilon^2, & \tau = 0, \\ 0, & \tau = 1, 2, \dots, \end{cases} \quad E(\epsilon_{t-1} y_{t-\tau}) = \begin{cases} (\phi_1 + \theta_1)\sigma_\epsilon^2, & \tau = 0, \\ \sigma_\epsilon^2, & \tau = 1, \\ 0, & \tau = 2, 3, \dots \end{cases}$$

Therefore, we obtain;

$$\gamma(0) = \phi_1 \gamma(1) + (1 + \phi_1 \theta_1 + \theta_1^2) \sigma_\epsilon^2,$$

$$\gamma(1) = \phi_1 \gamma(0) + \theta_1 \sigma_\epsilon^2,$$

$$\gamma(\tau) = \phi_1 \gamma(\tau - 1), \quad \tau = 2, 3, \dots$$

From the first two equations, $\gamma(0)$ and $\gamma(1)$ are computed by:

$$\begin{pmatrix} 1 & -\phi_1 \\ -\phi_1 & 1 \end{pmatrix} \begin{pmatrix} \gamma(0) \\ \gamma(1) \end{pmatrix} = \sigma_\epsilon^2 \begin{pmatrix} 1 + \phi_1 \theta_1 + \theta_1^2 \\ \theta_1 \end{pmatrix}$$

$$\begin{pmatrix} \gamma(0) \\ \gamma(1) \end{pmatrix} = \sigma_\epsilon^2 \begin{pmatrix} 1 & -\phi_1 \\ -\phi_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 + \phi_1 \theta_1 + \theta_1^2 \\ \theta_1 \end{pmatrix}$$

$$= \frac{\sigma_\epsilon^2}{1 - \phi_1^2} \begin{pmatrix} 1 & \phi_1 \\ \phi_1 & 1 \end{pmatrix} \begin{pmatrix} 1 + \phi_1\theta_1 + \theta_1^2 \\ \theta_1 \end{pmatrix} = \frac{\sigma_\epsilon^2}{1 - \phi_1^2} \begin{pmatrix} 1 + 2\phi_1\theta_1 + \theta_1^2 \\ (1 + \phi_1\theta_1)(\phi_1 + \theta_1) \end{pmatrix}.$$

Thus, the initial value of the autocorrelation coefficient is given by:

$$\rho(1) = \frac{(1 + \phi_1\theta_1)(\phi_1 + \theta_1)}{1 + 2\phi_1\theta_1 + \theta_1^2}.$$

We have:

$$\rho(\tau) = \phi_1\rho(\tau - 1).$$

ARMA(p, q) +drift:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}.$$

Mean of ARMA(p, q) Process: $\phi(L)y_t = \mu + \theta(L)\epsilon_t$,

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$.

$$y_t = \phi(L)^{-1} \mu + \phi(L)^{-1} \theta(L) \epsilon_t.$$

Therefore,

$$E(y_t) = \phi(L)^{-1} \mu + \phi(L)^{-1} \theta(L) E(\epsilon_t) = \phi(1)^{-1} \mu = \frac{\mu}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}.$$

7.6 ARIMA Model

Autoregressive Integrated Moving Average (ARIMA, 自己回帰和分移動平均) Model

ARIMA(p, d, q) Process

$$\phi(L)\Delta^d y_t = \theta(L)\epsilon_t,$$

where $\Delta^d y_t = \Delta^{d-1}(1-L)y_t = \Delta^{d-1}y_t - \Delta^{d-1}y_{t-1} = (1-L)^d y_t$ for $d = 1, 2, \dots$, and $\Delta^0 y_t = y_t$.

例 : ARIMA(0,1,0) Model

Consider the model: $\Delta y_t = y_t - y_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0, \sigma^2)$, $y_0 = 0$,

which is rewritten as: $y_t = \epsilon_t + \epsilon_{t-1} + \dots + \epsilon_1$.

$$E(y_t) = 0, \quad \gamma(0) = V(y_t) = \sigma^2 t, \quad \gamma(\tau) = \text{Cov}(y_t, y_{t-\tau}) = E(y_t y_{t-\tau}) = \sigma^2(t - \tau),$$

which implies that $\gamma(\tau)$ is time-dependent. $\implies y_t$ is not stationary.

$$\rho(\tau) = \frac{\text{Cov}(y_t, y_{t-\tau})}{\sqrt{V(y_t)} \sqrt{V(y_{t-\tau})}} = \frac{t - \tau}{\sqrt{t} \sqrt{t - \tau}} = \sqrt{\frac{t - \tau}{t}}.$$

That is, $\rho(\tau)$ gradually decreases with slow speed.

7.7 SARIMA Model

Seasonal ARIMA (SARIMA) Process:

1. SARIMA(p, d, q)

$$\phi(L)\Delta^d\Delta_s y_t = \theta(L)\epsilon_t,$$

where

$$\Delta_s y_t = (1 - L^s)y_t = y_t - y_{t-s}.$$

$s = 4$ when y_t denotes quarterly date and $s = 12$ when y_t represents monthly data.

7.8 Optimal Prediction

1. AR(p) Process: $y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t$

(a) Define:

$$E(y_{t+k}|Y_t) = y_{t+k|t},$$

where Y_t denotes all the information available at time t .

Taking the conditional expectation of $y_{t+k} = \phi_1 y_{t+k-1} + \cdots + \phi_p y_{t+k-p} + \epsilon_{t+k}$ on both sides,

$$y_{t+k|t} = \phi_1 y_{t+k-1|t} + \cdots + \phi_p y_{t+k-p|t},$$

where $y_{s|t} = y_s$ for $s \leq t$.

(b) Optimal prediction is given by solving the above differential equation.

2. MA(q) Process: $y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}$

(a) Let $\hat{\epsilon}_T, \hat{\epsilon}_{T-1}, \cdots, \hat{\epsilon}_1$ be the estimated errors.

(b) $y_{t+k} = \epsilon_{t+k} + \theta_1\epsilon_{t+k-1} + \cdots + \theta_q\epsilon_{t+k-q}$

(c) Therefore,

$$y_{t+k|t} = \epsilon_{t+k|t} + \theta_1\epsilon_{t+k-1|t} + \cdots + \theta_q\epsilon_{t+k-q|t},$$

where $\epsilon_{s|t} = 0$ for $s > t$ and $\epsilon_{s|t} = \hat{\epsilon}_s$ for $s \leq t$.

3. ARMA(p, q) Process: $y_t = \phi_1y_{t-1} + \cdots + \phi_py_{t-p} + \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}$

(a) $y_{t+k} = \phi_1y_{t+k-1} + \cdots + \phi_py_{t+k-p} + \epsilon_{t+k} + \theta_1\epsilon_{t+k-1} + \cdots + \theta_q\epsilon_{t+k-q}$

(b) Optimal prediction is:

$$y_{t+k|t} = \phi_1 y_{t+k-1|t} + \cdots + \phi_p y_{t+k-p|t} + \epsilon_{t+k|t} + \theta_1 \epsilon_{t+k-1|t} + \cdots + \theta_q \epsilon_{t+k-q|t},$$

where $y_{s|t} = y_s$ and $\epsilon_{s|t} = \hat{\epsilon}_s$ for $s \leq t$, and $\epsilon_{s|t} = 0$ for $s > t$.

7.9 Identification (識別, または, 同定)

1. Based on AIC or SBIC given d, s , we obtain p, q .

(a) AIC (Akaike's Information Criterion, 赤池の情報量基準)

$$\text{AIC} = -2 \log(\text{likelihood}) + 2k,$$

where $k = p + q$, which is the number of parameters estimated.

(b) SBIC (Shwarz's Bayesian Information Criterion)

$$\text{SBIC} = -2 \log(\text{likelihood}) + k \log T,$$

where T denotes the number of observations.

2. From the sample autocorrelation coefficient function $\hat{\rho}(k)$ and the partial autocorrelation coefficient function $\hat{\phi}_{k,k}$ for $k = 1, 2, \dots$, we obtain p, d, q, s .

	AR(p) Process	MA(q) Process
Autocorrelation Function	Gradually decreasing	$\rho(k) = 0,$ $k = q + 1, q + 2, \dots$
Partial Autocorrelation Function	$\phi(k, k) = 0,$ $k = p + 1, p + 2, \dots$	Gradually decreasing

(a) Compute $\Delta_s y_t$ to remove seasonality.

Compute the autocovariance functions of $\Delta_s y_t$.

If the autocovariance functions have period s , we take $(1 - L^s)$, again.

(b) Determine the order of difference.

Compute the partial autocovariance functions every time.

If the autocovariance functions decrease as τ is large, go to the next step.

(c) Determine the order of AR terms (i.e., p).

Compute the partial autocovariance functions every time.

The partial autocovariance functions are close to zero after some τ , go to the next step.

(d) Determine the order of MA terms (i.e., q).

Compute the autocovariance functions every time.

If the autocovariance functions are randomly around zero, end of the procedure.

7.10 Example of SARIMA using Consumption Data

Construct SARIMA model using monthly and seasonally unadjusted consumption expenditure data and STATA12.

Estimation Period: Jan., 1970 — Dec., 2012 ($T = 516$)

```
. gen time=_n
```

```
. tsset time  
    time variable:  time, 1 to 516  
        delta:    1 unit
```

```
. corrgram expend
```

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial Autocor]		
1	0.8488	0.8499	373.88	0.0000		-----		-----		-----
2	0.8231	0.3858	726.18	0.0000		-----		-----		-----
3	0.8716	0.5266	1122	0.0000		-----		-----		-----
4	0.8706	0.4025	1517.6	0.0000		-----		-----		-----
5	0.8498	0.3447	1895.3	0.0000		-----		-----		-----
6	0.8085	0.0074	2237.9	0.0000		-----		-----		-----
7	0.8378	0.1528	2606.5	0.0000		-----		-----		-----

8	0.8460	0.1467	2983	0.0000	-----	-
9	0.8342	0.3006	3349.9	0.0000	-----	--
10	0.7735	-0.1518	3666	0.0000	-----	-
11	0.7852	-0.1185	3992.3	0.0000	-----	
12	0.9234	0.9442	4444.5	0.0000	-----	-----
13	0.7754	-0.5486	4764.1	0.0000	-----	----
14	0.7482	-0.3248	5062.1	0.0000	-----	--
15	0.7963	-0.2392	5400.5	0.0000	-----	-

. gen dexp=expnd-1.expnd
(1 missing value generated)

. corrgram dexp

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]		[Partial Autocor]			
1	-0.4316	-0.4329	96.485	0.0000	---		---		---	
2	-0.2546	-0.5441	130.13	0.0000	--		----		----	
3	0.1721	-0.4091	145.53	0.0000	-		----		----	
4	0.0667	-0.3459	147.85	0.0000			---		---	
5	0.0715	-0.0036	150.52	0.0000						
6	-0.2428	-0.1489	181.36	0.0000	-				-	
7	0.0711	-0.1400	184.01	0.0000					-	
8	0.0668	-0.2900	186.36	0.0000					--	
9	0.1704	0.1681	201.64	0.0000	-					
10	-0.2485	0.1306	234.21	0.0000	-					
11	-0.4293	-0.9305	331.56	0.0000	---		-----		-----	
12	0.9773	0.6768	837.12	0.0000			-----		-----	
13	-0.4152	0.3778	928.56	0.0000	---				---	
14	-0.2583	0.2688	964.03	0.0000	--				--	
15	0.1712	0.0406	979.63	0.0000	-				-	

```
. gen sdex=dexp-112.dexp
(13 missing values generated)
```

```
. corrgram sdex
```

LAG	AC	PAC	Q	Prob>Q	⁻¹ [Autocorrelation]	⁰ [Partial Autocor]
1	-0.4752	-0.4753	114.28	0.0000	---	---
2	-0.0244	-0.3235	114.58	0.0000		--
3	0.1163	-0.0759	121.46	0.0000		
4	-0.1246	-0.1365	129.37	0.0000		-
5	0.0341	-0.1016	129.96	0.0000		
6	-0.0151	-0.1136	130.08	0.0000		
7	-0.0395	-0.1413	130.88	0.0000		-
8	0.1123	0.0092	137.35	0.0000		
9	-0.0664	-0.0100	139.62	0.0000		
10	0.0168	0.0069	139.76	0.0000		
11	0.1642	0.2422	153.68	0.0000		
12	-0.3888	-0.2469	231.9	0.0000	---	-
13	0.2242	-0.1205	257.96	0.0000		
14	-0.0147	-0.0941	258.07	0.0000		
15	-0.0708	-0.0591	260.68	0.0000		

```
. arima sdex, ar(1,2) ma(1)
```

```
(setting optimization to BHHH)
```

```
Iteration 0: log likelihood = -5107.4608
```

```
Iteration 1: log likelihood = -5102.391
```

Iteration 2: log likelihood = -5099.9071
 Iteration 3: log likelihood = -5099.4216
 Iteration 4: log likelihood = -5099.2463
 (switching optimization to BFGS)
 Iteration 5: log likelihood = -5099.2361
 Iteration 6: log likelihood = -5099.2346
 Iteration 7: log likelihood = -5099.2346
 Iteration 8: log likelihood = -5099.2346

ARIMA regression

Sample: 14 - 516

Log likelihood = -5099.235

Number of obs = 503
 Wald chi2(3) = 973.93
 Prob > chi2 = 0.0000

	sdex	Coef.	OPG Std. Err.	z	P> z	[95% Conf. Interval]	
sdex	_cons	-15.64573	59.17574	-0.26	0.791	-131.628	100.3366
ARMA	ar						
	L1.	.1271774	.0581883	2.19	0.029	.0131304	.2412244
	L2.	.1009983	.053626	1.88	0.060	-.0041068	.2061034
	ma						
	L1.	-.8343264	.0419364	-19.90	0.000	-.9165202	-.7521326
	/sigma	6111.128	139.0105	43.96	0.000	5838.673	6383.584

Note: The test of the variance against zero is one sided, and the two-sided confidence interval is truncated at zero.

. estat ic

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	503	.	-5099.235	5	10208.47	10229.57

Note: N=Obs used in calculating BIC; see [R] BIC note

. predict resid, r
(13 missing values generated)

. corrgram resid

LAG	AC	PAC	Q	Prob>Q	⁻¹ [Autocorrelation]	⁰ [Partial Autocor]	¹
1	-0.0132	-0.0132	.08814	0.7666			
2	-0.0095	-0.0097	.1341	0.9351			
3	0.1248	0.1246	8.0433	0.0451			
4	-0.0644	-0.0624	10.154	0.0379			
5	-0.0001	0.0011	10.154	0.0710			
6	-0.0138	-0.0309	10.252	0.1144			
7	-0.0032	0.0126	10.257	0.1745			
8	0.0958	0.0938	14.97	0.0597			
9	-0.0317	-0.0255	15.487	0.0784			

10	0.0126	0.0112	15.569	0.1127
11	-0.0053	-0.0305	15.583	0.1573
12	-0.3773	-0.3837	89.235	0.0000
13	0.0408	0.0258	90.098	0.0000
14	-0.0233	-0.0307	90.381	0.0000
15	-0.0911	-0.0059	94.703	0.0000

---		---	
-----	--	-----	--