FORECASTING
STAT0010

Alexandros Beskos

a.beskos@ucl.ac.uk

'Lecture 6' Outline

1. ARIMA and SARMA models
2. SARIMA models
3. Identification (so far)
4. Partial Autocorrelation Function

Definition 1 (seasonal moving average $SMA(Q)_s$ model)

$\{Y_t\}$ is a seasonal moving average process of order Q, with period s, written as $SMA(Q)_s$, if

$$Y_t = \epsilon_t - \Theta_1 \epsilon_{t-s} - \Theta_2 \epsilon_{t-2s} - \ldots - \Theta_Q \epsilon_{t-Qs}\,, \quad s \geq 2$$

Definition 2 (seasonal autoregressive $SAR(P)_s$ model)

$\{Y_t\}$ is a seasonal autoregressive process of order P, with period s, written as $SAR(P)_s$, if

$$Y_t = \Phi_1 Y_{t-s} + \Phi_2 Y_{t-2s} + \ldots + \Phi_P Y_{t-Ps} + \epsilon_t\,, \quad s \geq 2$$

Definition 3 ($SARMA(p, q) \times (P, Q)_s$ model)

$\{Y_t\}$ *is a multiplicative seasonal autoregressive, moving average process of order* $(p, q) \times (P, Q)_s$ *, with period* s*, written as* $SARMA(p, q) \times (P, Q)_s$ *, if*

$$\phi(B)\, \Phi(B)\, Y_t = \theta(B)\, \Theta(B)\, \epsilon_t$$

Definition 4 ($ARIMA(p, d, q)$ model)

Let the process $\{\nabla^d Y_t\}$ *be an* $ARMA(p, q)$ *process of order* (p, q)*. Then,* $\{Y_t\}$ *is an integrated autoregressive, moving average process of order* (p, d, q)*, written* $ARIMA(p, d, q)$*, with model equation:*

$$\phi(B)(1 - B)^d Y_t = \theta(B)\epsilon_t \, .$$

Question 5

(How) can we combine $ARIMA$ *and* $SARMA$*?*

Definition 6

Let $\{Y_t\}$ *be some process. Then (recall that) the difference operator* ∇ *is defined by*

$$\nabla Y_t = Y_t - Y_{t-1}\,.$$

The d*th difference operator* ∇^d *is defined as*

$$\nabla^d Y_t = (1 - B)^d Y_t\,.$$

Definition 7

Let $\{Y_t\}$ *be some process. Then the seasonal difference operator* ∇_s *is defined by*

$$\nabla_s Y_t = Y_t - Y_{t-s}\,.$$

The D*th seasonal difference operator* ∇^D *is defined as*

$$\nabla_s^D Y_t = (1 - B^s)^D\,.$$

Remark 8

Note $\nabla_s^2 = \nabla_s \nabla_s$, $\nabla_s^3 = \nabla_s \nabla_s \nabla_s$, *etc.*

Definition 9 (Integrated, seasonal, autoregressive, moving average process: *SARIMA*)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$ and let the process $\{\nabla^d \nabla_s^D Y_t\}$ be a $SARMA(p, q) \times (P, Q)_s$ process of order $(p, q) \times (P, Q)_s$ with period s. Then, $\{Y_t\}$ is an integrated, seasonal, autoregressive, moving average process of order $(p, d, q) \times (P, D, Q)_s$, with period s, written $SARIMA(p, d, q) \times (P, D, Q)_s$, with model equation:

$$\phi(B)\,\Phi(B)\,\nabla^d\,\nabla_s^D\,Y_t = \theta(B)\,\Theta(B)\,\epsilon_t$$

Remark 10

If $\{Y_t\}$ has to be (non-seasonally) differenced d-many times and seasonally differenced D-many times before it is a stationary $SARMA(p, q) \times (P, Q)_s$ process, then $\{Y_t\}$ is a $SARIMA(p, d, q) \times (P, D, Q)_s$ process.

Example 11 (on board)

- $SARIMA(1, 0, 0) \times (0, 1, 1)_{12}$
- $SARIMA(0, 1, 0) \times (1, 0, 1)_4$

Example 12

Just by plotting at the time series, the following data looks suspiciously like $SARIMA(p,1,q)\times(P,1,Q)_{12}$.

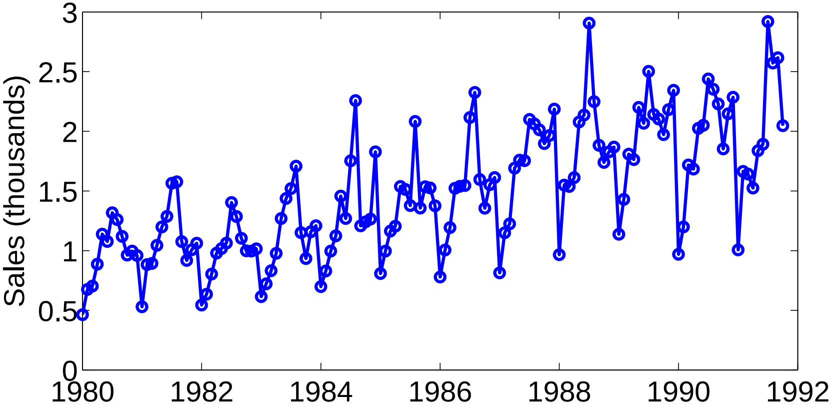


Figure: Australian red wine sales, Jan. '80 — Oct. '91

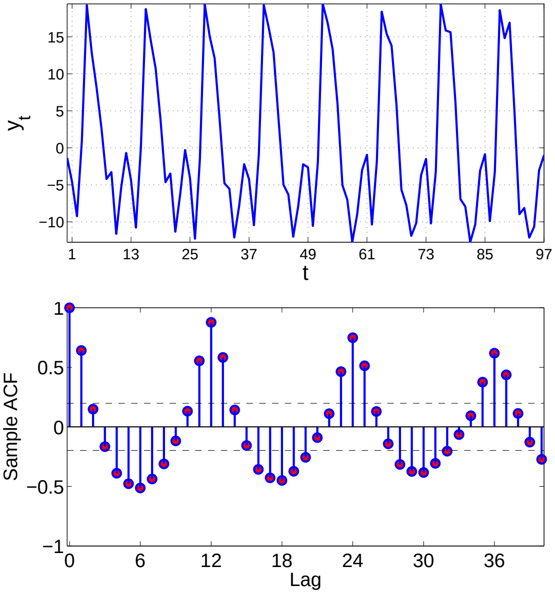
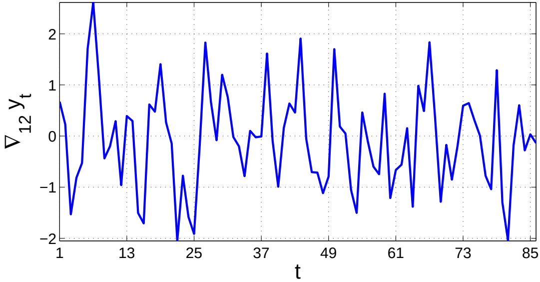


Figure: Simulated $SARIMA(0,0,1) \times (0,1,0)_{12}$, with ACF



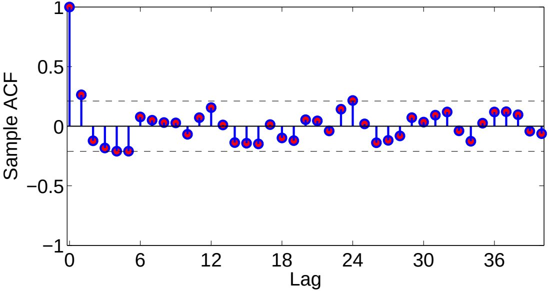


Figure: Seasonally differenced $SARIMA(0,0,1) \times (0,1,0)_{12}$, with ACF

The Box-Jenkins methodology for forecasting

1. **Model identification**
    - Look at data. Compute sample ACF. Try to deduce whether model is $AR(p)$, $MA(q)$, $ARMA(p,q)$, $ARIMA(p,d,q)$, $SAR(P)_s$, $SMA(Q)_s$, $SARMA(p,q) \times (P,Q)_s$, $SARIMA(p,d,q) \times (P,D,Q)_s$; decide on reasonable values for p, d, q, P, D, Q, s.
2. **Parameter estimation**
    - Using the model and values of (the model orders) p, q, etc. from the first step, estimate the unknown parameters, $\mu, \phi_1, \phi_2, \ldots, \phi_p$, $\theta_1, \ldots, \theta_q$, $\Phi_1, \Phi_2, \ldots$, $\Theta_1, \Theta_2, \ldots$, d, D, s, etc.
3. **Verification**
   Check model obtained from 1 & 2
    - Good? Goto 4
    - Bad? Goto 1 & decide on new model
4. **Forecasting**

model identification with the ACF

Model	ACF
$AR(1)$	$\rho(k) = \phi_1^k$: exponential decay for $0 < \phi_1 < 1$ (alternating exponential decay if $-1 < \phi_1 < 0$)
$AR(p)$	exponential decay or damped sinusoid
$SAR(1)_s$	zeros at $k \neq \ell s$; non-zeros at lags ℓs, $\ell \in \mathbb{Z}$ decay exponentially
$MA(1)$	$\rho(1) = \frac{-\theta_1}{1+\theta_1^2}$: 'spike' at lag 1, then 0 for lags ≥ 2 (spike is positive if $\theta_1 < 0$ and negative if $\theta_1 > 0$)
$MA(q)$	spikes at lags 1 to q and 0 for lags $\geq q+1$
$SMA(1)_s$	$\rho(ks) = \frac{-\Theta_1}{1+\Theta_1^2}$: 'spike' at lag s, and 0 otherwise
$ARMA(p, q)$	exponential decay or damped sinusoid (for lags $> q$)
$SARMA$	'periodically extended' version of non-seasonal case

Question 13

(How) can we design something like the ACF which cuts-off at pth lag for an $AR(p)$ process (in the same way the ACF cuts-off at qth lag for an $MA(q)$ process)?

Remark 14

$AR(p)$ processes do not have an ACF that cuts-off at lag p because future values of series depend (indirectly) on **all** *previous values.*

Example 15

Recall $AR(1)$: $Y_t = \phi_1 Y_{t-1} + \epsilon_t$, with ACF $\rho(k) = \phi_1^{|k|}$.

Using successive substitution:

$$\begin{aligned} Y_2 &= \phi_1 Y_1 + \epsilon_2 \\ Y_3 &= \phi_1 Y_2 + \epsilon_3 = \phi_1^2 Y_1 + \phi_1 \epsilon_2 + \epsilon_3 \,, \end{aligned}$$

i.e. Y_3 depends on Y_1. In general, Y_t depends on $Y_{t-1}, Y_{t-2}, \ldots$ I.e. Y_t depends on **all** previous values of the process $\{Y_t\}$.

- Consider representing an $AR(1)$ process as a linear combination of previous two values:

$$Y_t = \phi_{2,1} Y_{t-1} + \phi_{2,2} Y_{t-2} + \epsilon_t \,, \quad \text{for some } \phi_{2,1}, \phi_{2,2} \in \mathbb{R} \,.$$

  Then, $\phi_{2,2}$ represents any linear dependence that Y_t has on Y_{t-2} which is not accounted for by Y_{t-1}.

- For an $AR(1)$ process $Y_t = \phi_1 Y_{t-1} + \epsilon_t$, we have, by definition, that $\phi_{2,1} = \phi_1$ and $\underline{\underline{\phi_{2,2} = 0}}$.

- If we now consider representing this $AR(1)$ by:

$$Y_t = \phi_{1,1} Y_{t-1} + \epsilon_t \quad \phi_{1,1} \in \mathbb{R} \,,$$

  then, again, $\underline{\underline{\phi_{1,1} = \phi_1 \neq 0}}$.

Remark 16

The numbers $\phi_{1,1}$ and $\phi_{2,2}$ are the first two partial autocorrelation coefficients (PACF). Note that, for the $AR(1)$ process, the PACF cuts off at lag 2.

Definition 17

Consider (the problem of estimating) the coefficients $\phi_{k,j} \in \mathbb{R}$, *where*

$$\begin{aligned} Y_t &= \phi_{1,1} Y_{t-1} + \epsilon_t \\ Y_t &= \phi_{2,1} Y_{t-1} + \phi_{2,2} Y_{t-2} + \epsilon_t \\ Y_t &= \phi_{3,1} Y_{t-1} + \phi_{3,2} Y_{t-2} + \phi_{3,3} Y_{t-3} + \epsilon_t \\ &\vdots \\ Y_t &= \phi_{k,1} Y_{t-1} + \phi_{k,2} Y_{t-2} + \ldots + \phi_{k,k} Y_{t-k} + \epsilon_t \end{aligned}$$

The sequence $\{\phi_{k,k}\} = \{\phi_{1,1}, \phi_{2,2}, \phi_{3,3}, \ldots\}$ *is called the partial autocorrelation function (PACF) coefficients.*

Example 18

For $AR(1)$, *by definition,* $\phi_{1,1} \neq 0$, *and* $\phi_{2,2}, \phi_{3,3}, \ldots = 0$.

Example 19

For $AR(2)$, *by definition,* $\phi_{1,1}, \phi_{2,2} \neq 0$, *and* $\phi_{3,3}, \phi_{4,4}, \ldots = 0$.

Remark 20

For $AR(p)$, by definition:

$$\begin{aligned} \phi_{k,k} &\neq 0, \quad \textit{for } k = 1, \ldots p \\ \phi_{k,k} &= 0, \quad \textit{for } k \geq p+1 \end{aligned}$$

Recall

$$Y_t = \phi_{k,1} Y_{t-1} + \phi_{k,2} Y_{t-2} + \ldots + \phi_{k,k} Y_{t-k} + \epsilon_t \tag{1}$$

Multiply both sides by Y_{t-j}, for $j \geq 0$:

$$Y_{t-j} Y_t = \phi_{k,1} Y_{t-j} Y_{t-1} + \phi_{k,2} Y_{t-j} Y_{t-2} + \ldots + \phi_{k,k} Y_{t-j} Y_{t-k} + Y_{t-j} \epsilon_t$$

Take expectations of both sides:

$$\gamma(j) = \phi_{k,1} \gamma(j-1) + \phi_{k,2} \gamma(j-2) + \ldots + \phi_{k,k} \gamma(j-k)$$

Divide by $\gamma(0)$:

$$\rho(j) = \phi_{k,1} \rho(j-1) + \phi_{k,2} \rho(j-2) + \ldots + \phi_{k,k} \rho(j-k)$$

$$\rho(j) = \phi_{k,1}\rho(j-1) + \phi_{k,2}\rho(j-2) + \ldots + \phi_{k,k}\rho(j-k)$$

For $j = 1$:

$$\rho(1) = \phi_{k,1}\rho(0) + \phi_{k,2}\rho(1) + \phi_{k,3}\rho(2) + \phi_{k,4}\rho(3) + \ldots + \phi_{k,k}\rho(k-1)$$

For $j = 2$:

$$\rho(2) = \phi_{k,1}\rho(1) + \phi_{k,2}\rho(0) + \phi_{k,3}\rho(1) + \phi_{k,4}\rho(2) + \ldots + \phi_{k,k}\rho(k-2)$$

For $j = 3$:

$$\rho(3) = \phi_{k,1}\rho(2) + \phi_{k,2}\rho(1) + \phi_{k,3}\rho(0) + \phi_{k,4}\rho(1) + \ldots + \phi_{k,k}\rho(k-3)$$

$$\vdots$$

For $j = k$:

$$\rho(k) = \phi_{k,1}\rho(k-1) + \phi_{k,2}\rho(k-2) + \phi_{k,3}\rho(k-3) + \phi_{k,4}\rho(k-4) + \ldots + \phi_{k,k}\rho(0)$$

Can be written as the Yule-Walker equations:

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \rho(4) \\ \vdots \\ \rho(k) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \rho(3) & \dots & \rho(k-1) \\ \rho(1) & 1 & \rho(1) & \rho(2) & \dots & \rho(k-2) \\ \rho(2) & \rho(1) & 1 & \rho(1) & \dots & \rho(k-3) \\ \rho(3) & \rho(2) & \rho(1) & 1 & \dots & \rho(k-4) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \rho(k-4) & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k,1} \\ \phi_{k,2} \\ \phi_{k,3} \\ \phi_{k,4} \\ \vdots \\ \phi_{k,k} \end{bmatrix}$$

I.e.

$$\boldsymbol{\rho}_k = \boldsymbol{R}_k \boldsymbol{\phi}_k \, ,$$

where $\boldsymbol{\rho}_k, \boldsymbol{\phi}_k \in \mathbb{R}^k$, and $\boldsymbol{R}_k \in \mathbb{R}^{k \times k}$. It can be shown ($\boldsymbol{R}_k$ is semi-positive definite) that $\boldsymbol{R}$ is invertible. Hence:

$$\boldsymbol{\phi}_k = \boldsymbol{R}_k^{-1} \boldsymbol{\rho}_k \, .$$

Note that $\boldsymbol{R}_k$ and $\boldsymbol{\rho}_k$ only contain ACF coefficients. Therefore we can compute the PACF coefficients from the ACF.

$$\boldsymbol{\phi}_k = \boldsymbol{R}_k^{-1}\boldsymbol{\rho}_k\,.$$

Example 21

$\boldsymbol{\phi}_1 = \phi_{1,1} = \boldsymbol{R}_1^{-1}\boldsymbol{\phi}_1 = 1\rho(1)$. *Hence*

$$\boxed{\phi_{1,1} = \rho(1)}$$

Example 22

$$\begin{aligned}
\boldsymbol{\phi}_2 = \begin{bmatrix} \phi_{2,1} \\ \phi_{2,2} \end{bmatrix} &= \begin{bmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho(1) \\ \rho(2) \end{bmatrix} \\
&= \frac{1}{1-\rho(1)^2} \begin{bmatrix} 1 & -\rho(1) \\ -\rho(1) & 1 \end{bmatrix} \begin{bmatrix} \rho(1) \\ \rho(2) \end{bmatrix} \\
&= \frac{1}{1-\rho(1)^2} \begin{bmatrix} \rho(1)(1-\rho(2)) \\ \rho(2)-\rho(1)^2 \end{bmatrix}
\end{aligned}$$

Hence

$$\boxed{\phi_{2,2} = \frac{\rho(2)-\rho(1)^2}{1-\rho(1)^2}}$$

$$\begin{aligned}\phi_{1,1} &= \rho(1)\,,\\ \phi_{2,2} &= \frac{\rho(2)-\rho(1)^2}{1-\rho(1)^2}\,.\end{aligned}$$

Example 23

For $AR(1)$, recall $\rho(k) = \phi_1^{|k|}$. Then

$$\phi_{1,1} = \rho(1) = \phi_1\,,$$

and

$$\phi_{2,2} = \frac{\phi_1^2 - \phi_1^2}{1-\phi_1^2} = 0\,.$$

I.e., 'cut-off' at lag 2.

We can find $\phi_{3,3}$, $\phi_{4,4}$, etc. in a similar way by solving higher order sets of Yule-Walker equations, e.g. $\boldsymbol{\phi}_3 = \boldsymbol{R}_3^{-1}\boldsymbol{\rho}_3$, $\boldsymbol{\phi}_4 = \boldsymbol{R}_4^{-1}\boldsymbol{\rho}_4$, etc.

However, (perhaps unsurprisingly?) there is a more efficient way...

Theorem 24 (Durbin-Levinson)

The PACF coefficients can be computed via:

$$\phi_{k,k} = \frac{\rho(k) - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho(k-j)}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho(j)}$$

where $\phi_{k,j} = \phi_{k-1,j} - \phi_{k,k}\phi_{k-1,k-j}$, *for* $j = 1, 2, \ldots, k-1$.

Example 25

A time series model has $\rho(1) = 2/5$, $\rho(2) = -1/20$, $\rho(3) = -1/8$. *Find PACF at lags* $1, 2, 3$.

For $k = 1$: $\phi_{1,1} = \rho(1) = \underline{\underline{2/5}}$. For $k = 2$:

$$\phi_{2,2} = \frac{\rho(2) - \sum_{j=1}^{1} \phi_{1,j}\rho(2-j)}{1 - \sum_{j=1}^{1} \phi_{1,j}\rho(j)} = \frac{\rho(2) - \phi_{1,1}\rho(1)}{1 - \phi_{1,1}\rho(1)} = \frac{-1/20 - (2/5)^2}{1 - (2.5)^2}$$

I.e. $\phi_{2,2} = \underline{\underline{-1/4}}$.

Theorem 24 (Durbin-Levinson)

The PACF coefficients can be computed via:

$$\phi_{k,k} = \frac{\rho(k) - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho(k-j)}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho(j)}$$

where $\phi_{k,j} = \phi_{k-1,j} - \phi_{k,k}\phi_{k-1,k-j}$, *for* $j = 1, 2, \ldots, k-1$.

Example 25

A time series model has $\rho(1) = 2/5$, $\rho(2) = -1/20$, $\rho(3) = -1/8$. *Find PACF at lags* $1, 2, 3$.

For $k = 3$:

$$\phi_{3,3} = \frac{\rho(3) - \sum_{j=1}^{2} \phi_{2,j}\rho(3-j)}{1 - \sum_{j=1}^{2} \phi_{2,j}\rho(j)} = \frac{\rho(3) - (\phi_{2,1}\rho(2) + \phi_{2,2}\rho(1))}{1 - (\phi_{2,1}\rho(1) + \phi_{2,2}\rho(2))}$$

where $\phi_{2,1} = \phi_{1,1} - \phi_{2,2}\phi_{1,1} = 4/10 - (1/4)4/10 = 1/2$. I.e.

$$\phi_{3,3} = \underline{\underline{0}}.$$

In practice, the **sample** PACF coefficients $\hat{\phi}_{k,k}$ can be computed from the **sample** ACF coefficients $\hat{\rho}(k)$, using Yule-Walker equations $\hat{\boldsymbol{\phi}}_k = \hat{\boldsymbol{R}}_k^{-1} \hat{\boldsymbol{\rho}}_k$, or Durbin Levinson.

model identification with the ACF and PACF

Model	ACF	PACF
$AR(1)$	exponential decay	spike lag 1, then 0
$AR(p)$	exponential decay or damped sinusoid	spikes lags 1 to p, then 0
$MA(1)$	spike lag 1, then 0	exponential decay
$MA(q)$	spikes lags 1 to q, then 0	exponential decay or damped sinusoid
$ARMA(p, q)$	exponential decay or damped sinusoid (for lags $> q$)	exponential decay or damped sinusoid (for lags $> p$)

Remark 26

We can test whether PACF coefficients are significantly different from (a 'genuine') zero by using the 95% *confidence interval* $(-1.96/\sqrt{T}, 1.96/\sqrt{T})$ *(c.f. white noise remark, lecture 1 and ACF remark, lecture 4).*

Example 27 ($T = 120 \Rightarrow$ confidence interval: $\approx (\pm 0.18)$)

k	0	1	2	3	4	5	6
Sample ACF	1	−0.52	−0.04	0.13	−0.09	−0.01	0.10
Sample PACF	1	−0.52	−0.43	−0.20	−0.19	−0.21	−0.06

Example 28 ($T = 120 \Rightarrow$ confidence interval: $\approx (\pm 0.18)$)

k	0	1	2	3	4	5	6
Sample ACF	1	0.44	−0.18	−0.31	0.01	0.25	0.13
Sample PACF	1	0.44	−0.46	0.02	0.19	0.04	−0.07

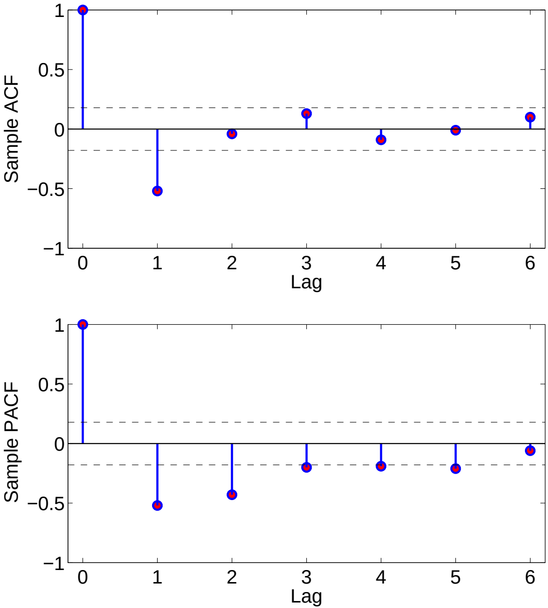


Figure: Example 27: Identify this model?

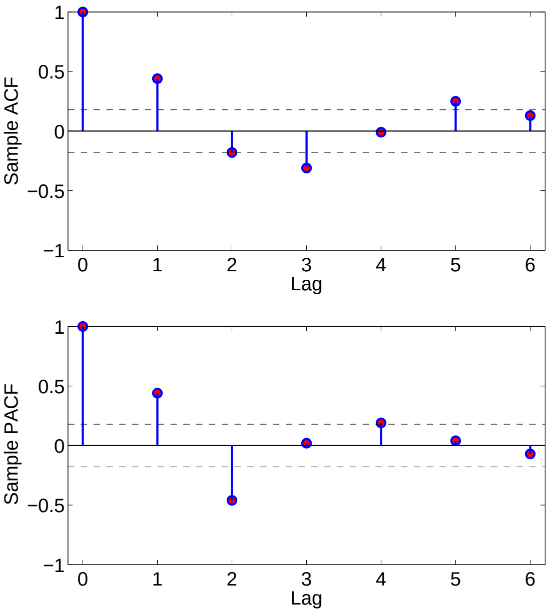


Figure: Example 28: Identify this model?