FORECASTING
STAT0010

Alexandros Beskos

a.beskos@ucl.ac.uk

'Lecture 3' Outline

1. Autoregressive processes
   - $AR(1)$
   - backshift operator
   - $AR(2)$ process
   - $AR(2)$ process example
   - Yule-Walker
2. ARMA models
   - definitions
   - $ARMA(p, q)$ and zero-mean
   - $ARMA(1, 1)$ and ACF via Yule-Walker
   - Box-Jenkins

Definition 1 (Autoregressive process)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$. Then, $\{Y_t\}$ is an autoregressive process of order p, written $AR(p)$, if

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t \, ,$$

Remark 2 ($AR(1)$ is $MA(\infty)$!)

$AR(1)$ can be written as:

$$Y_t = \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j} \, , \qquad |\phi_1| < 1$$

Corollary 3 (mean and γ for $AR(1)$)

Let $\{Y_t\}$ be $AR(1)$, with $|\phi_1| < 1$. Then,

1. $\mathbb{E}(Y_t) = 0$
2. $\gamma(k) = \dfrac{\sigma^2 \phi_1^k}{1 - \phi_1^2}$

Definition 4 (Backshift operator)

Let $\{Y_t\}$ *be some time series. Then the backshift operator (a.k.a lag operator)* B *is defined by*

$$B^k Y_t = Y_{t-k}$$

Example 5

Consider $AR(1)$ *process:* $Y_t = \phi_1 Y_{t-1} + \epsilon_t$. *In terms of backshift operator:*

$$\begin{aligned} Y_t &= \phi_1 B Y_t + \epsilon_t \\ Y_t - \phi_1 B Y_t &= \epsilon_t \\ (1 - \phi_1 B) Y_t &= \epsilon_t \\ Y_t &= (1 - \phi_1 B)^{-1} \epsilon_t = \sum_{j=0}^{\infty} (\phi_1 B)^j \epsilon_t = \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j} \end{aligned}$$

provided $|\phi_1| < 1$. *N.b.* $\phi(x) := (1 - \phi_1 x)$ *is called the* $AR(1)$ *charateristic polynomial/equation. Hence* $AR(1)$ *is stationary iff the root* $(1/\phi_1)$ *of* $\phi(x)$ *lies outside unit circle (i.e. if* $|1/\phi_1| > 1$*).*

Example 6 ($AR(2)$ process)

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t \tag{1}$$

Writing in terms of backshift operator:

$$\begin{aligned} Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} &= \epsilon_t \\ (1 - \phi_1 B - \phi_2 B^2) Y_t &= \epsilon_t \end{aligned}$$

$$\phi(B) Y_t = \epsilon_t$$

Now, $\{Y_t\}$ is stationary iff roots of $\phi(B) = (1 - \phi_1 B - \phi_2 B^2)$ lie outside unit circle. It can be shown that this is true iff:

$$\phi_1 + \phi_2 < 1\,, \quad \phi_2 - \phi_1 < 1\,, \quad |\phi_2| < 1 \tag{2}$$

Note also, if $\mu := \mathbb{E}(Y_t)$, then take $\mathbb{E}$ of (1):

$$\mu = (\phi_1 + \phi_2)\mu$$

But (2) $\Rightarrow (\phi_1 + \phi_2) \neq 1$. Hence $\mu = 0$.

Example 7 ($AR(2)$ process example)

Consider the case $\phi_1 = 1/6, \phi_2 = -1/6$*:*

$$Y_t = -\frac{1}{6}Y_{t-1} + \frac{1}{6}Y_{t-2} + \epsilon_t\,.$$

Is this process stationary?

Writing in terms of backshift operator:

$$\begin{aligned} Y_t + \frac{1}{6}Y_{t-1} - \frac{1}{6}Y_{t-2} &= \epsilon_t \\ (1 + \frac{1}{6}B - \frac{1}{6}B^2)Y_t &= \epsilon_t \end{aligned}$$

Factorise AR characteristic polynomial:

$$1 + \frac{1}{6}B - \frac{1}{6}B^2 = -\frac{1}{6}(B-2)(B+3)\,.$$

$\Rightarrow$ roots at 2 and -3. Hence, Y_t is stationary. Also note that

$$\phi_1 + \phi_2 < 1\,, \quad \phi_2 - \phi_1 < 1\,, \quad |\phi_2| < 1$$

are satisfied.

Definition 8 (causality)

A time series $\{Y_t\}$ is called causal if it, and any noise terms $\{\epsilon_t\}$, satisfy:

$$Y_{t-k} \perp \epsilon_t \,, \quad \forall k > 0$$

Remark 9

Causality means that 'future noise' is independent of present or past values of the time series. This should, intuitively, make sense. We will restrict our attention to causal models in this course.

Lemma 10

Let $\{Y_t\}$ be a causal time series and $\{\epsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$, then $\mathbb{E}(Y_{t-k}\epsilon_t) = 0 \,, \forall k > 0$.

Proof $\forall k > 0$:

$$\begin{aligned} 0 &= \mathrm{cov}(Y_{t-k}, \epsilon_t)\,, \quad \text{since } \{Y_t\} \text{ causal } \Rightarrow Y_{t-k} \perp \epsilon_t \\ &= \mathbb{E}(Y_{t-k}\epsilon_t) - \mathbb{E}(Y_{t-k})\cancelto{0}{\mathbb{E}(\epsilon_t)} \qquad \blacksquare \end{aligned}$$

Remark 11

Assuming that $\{Y_t\}$ is stationary, the ACF of an AR process can be found via the Yule-Walker equations.

Example 12 ($AR(2)$ process)

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$$

Multiply both sides by Y_{t-k} and take expectations (n.b. $\{Y_t\}$ causal):

$$\mathbb{E}(Y_{t-k} Y_t) = \phi_1 \mathbb{E}(Y_{t-k} Y_{t-1}) + \phi_2 \mathbb{E}(Y_{t-k} Y_{t-2}) + \underbrace{\cancel{\mathbb{E}(Y_{t-k}\epsilon_t)}}_{0,\, k>0}$$

Now, from $AR(2)$ Example 6, we showed $\mathbb{E}(Y_t) = 0$. Hence,

$$\text{cov}(Y_t, Y_{t-k}) = \mathbb{E}(Y_t Y_{t-k}) - \underbrace{\cancel{\mathbb{E}(Y_t)\mathbb{E}(Y_{t-k})}}_{0},$$

I.e. for $k > 0$:

$$\begin{aligned} \gamma(k) &= \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) \\ \rho(k) &= \phi_1 \rho(k-1) + \phi_2 \rho(k-2), \qquad \text{divide by } \gamma(0) \end{aligned}$$

$$\rho(k) = \phi_1\rho(k-1) + \phi_2\rho(k-2)$$

E.g., put $k = 1$:

$$\rho(1) = \phi_1\rho(0) + \phi_2\rho(-1)$$

But, $\rho(0) = 1$, and $\rho(-k) = \rho(k) \Rightarrow \rho(-1) = \rho(1)$:

$$\rho(1) = \phi_1 + \phi_2\rho(1) \Rightarrow \rho(1) = \frac{\phi_1}{1-\phi_2}$$

We can continue by putting $k = 2$ to get:

$$\rho(2) = \phi_1\rho(1) + \phi_2\rho(0)$$

I.e.

$$\rho(2) = \frac{\phi_1^2}{1-\phi_2} + \phi_2\,,$$

and so on. (Exercise: compute $\rho(3)$).
For an $AR(p)$ process, it can be shown that real roots $\Rightarrow$ exponentially decaying ACF and non-real roots $\Rightarrow$ an ACF with exponential decay multiplied by sum of sinusoids.

Remark 13 (Yule-Walker for $AR(p)$)

In general, the ACF of $AR(p)$: $Y_t = \epsilon_t + \sum_{j=1}^{p} \phi_j Y_{t-j}$ can also be found by Yule-Walker.

Assuming stationarity, multiply by Y_{t-k}, take $\mathbb{E}$, and divide by $\gamma(0)$:

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2) + \cdots + \phi_p \rho(k-p)$$

We can then solve these difference equations with standard methods (**but this is beyond scope of this course!**).

Also note: (recall matrix algebra in lecture 1) these can be written as:

$$\begin{pmatrix} \rho(1) \\ \vdots \\ \rho(p) \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & \dots & \rho(p-1) \\ \rho(1) & 1 & \dots & \rho(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(p-1) & \rho(p-2) & \dots & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix}$$

I.e.

$$\boldsymbol{\rho} = \boldsymbol{\Gamma} \boldsymbol{\phi} \,,$$

with

$$\boldsymbol{\rho} := (\rho(k))_{k=1}^{p} \,, \quad \boldsymbol{\phi} := (\phi_j)_{j=1}^{p} \,, \quad \boldsymbol{\Gamma} := (\rho(k-j))_{k,j=1}^{p}$$

Definition 14 (Moving average $MA(q)$ process)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$. Then, $\{Y_t\}$ is a moving average process of order q, written $MA(q)$, if

$$Y_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}$$

Definition 15 (Autoregressive $AR(p)$ process)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$. Then, $\{Y_t\}$ is an autoregressive process of order p, written $AR(p)$, if

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t \, ,$$

Definition 16 (Autoregressive, moving average $ARMA(p, q)$ process)

Let $\{\epsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$. Then, $\{Y_t\}$ is an autoregressive, moving average process of order (p, q), written $ARMA(p, q)$, if it contains p-many AR terms and q-many MA terms:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}$$

Remark 17

Again, as in purely $AR(p)$ case, $\{Y_t\}$ is stationary iff all roots of AR characteristic equation are strictly greater than one.

Lemma 18

The $ARMA(p, q)$ process:

$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}$

has zero-mean.

Proof

$$Y_t = \sum_{j=0}^{\infty} \phi_j Y_{t-j} + \epsilon_t - \sum_{k=0}^{\infty} \theta_k \epsilon_{t-k} \tag{3}$$

Let $\mathbb{E}(Y_t) = \mu$. Take $\mathbb{E}$ of both sides of (3):

$$\begin{aligned} \mathbb{E}(Y_t) &= \sum_{j=0}^{\infty} \phi_j \mathbb{E} Y_{t-j} \\ \mu &= \mu \sum_{j=0}^{\infty} \phi_j \end{aligned}$$

$$\mu = \mu \sum_{j=0}^{\infty} \phi_j$$

Hence, $\sum_{j=1}^{p} \phi_j \neq 1 \Rightarrow \mu = 0$. Now, recall $\{Y_t\}$ stationary $\Rightarrow AR$ characteristic equation has roots outside unit circle, i.e. roots of:

$$1 - \sum_{j=1}^{p} \phi_j x^j = 0 \tag{4}$$

must have magnitude greater than one. But, if $\sum_{j=1}^{p} \phi_j = 1$, then (4) has a solution at $x = 1$. Therefore, we must have that $\mu = 0$. $\blacksquare$

Example 19 ($ARMA(1,1)$)

Consider the $ARMA(1,1)$ model.

$$Y_t = \phi_1 Y_{t-1} + \epsilon_t - \theta_1 \epsilon_{t-1} \tag{5}$$

We can, again, use Yule-Walker-like arguments to derive the ACF.

Multiply both sides of (5) by Y_{t-k} and take expectations:

$$\mathbb{E}(Y_t Y_{t-k}) = \phi_1 \mathbb{E}(Y_{t-1} Y_{t-k}) + \mathbb{E}(\epsilon_t Y_{t-k}) - \theta_1 \mathbb{E}(\epsilon_{t-1} Y_{t-k}) \tag{6}$$

Now, look at 2nd and 3rd terms on RHS. Shift time index in Equation (5) by $-k$ and multiply through by ϵ_t; then $\forall k \geq 0$:

$$\begin{aligned}
\mathbb{E}(\epsilon_t Y_{t-k}) &= \phi_1 \cancelto{0}{\mathbb{E}(\epsilon_t Y_{t-k-1})} + \mathbb{E}(\epsilon_t \epsilon_{t-k}) - \theta_1 \cancelto{0}{\mathbb{E}(\epsilon_t \epsilon_{t-k-1})} = \sigma^2 \delta_{0,k}\,, \\
\mathbb{E}(\epsilon_{t-1} Y_{t-k}) &= \phi_1 \mathbb{E}(\epsilon_{t-1} Y_{t-k-1}) + \mathbb{E}(\epsilon_{t-1} \epsilon_{t-k}) - \theta_1 \mathbb{E}(\epsilon_{t-1} \epsilon_{t-k-1}) \\
&= \phi_1 \sigma^2 \delta_{0,k} + \sigma^2 \delta_{1,k} - \theta_1 \sigma^2 \delta_{0,k} \\
&= \sigma^2 (\phi_1 - \theta_1) \delta_{0,k} + \sigma^2 \delta_{1,k}
\end{aligned}$$

Hence, substituting these into (6):

$$\gamma(k) = \phi_1 \gamma(k-1) + \sigma^2 \delta_{0,k} - \theta_1 \sigma^2 \left((\phi_1 - \theta_1) \delta_{0,k} + \delta_{1,k} \right)$$

$$\gamma(k) = \phi_1\gamma(k-1) + \sigma^2\delta_{0,k} - \theta_1\sigma^2\big((\phi_1-\theta_1)\delta_{0,k} + \delta_{1,k}\big)$$

In particular, at $k = 0, 1, \dots$:

$$\begin{aligned}
\gamma(0) &= \phi_1\gamma(1) + \sigma^2(1-\theta_1\phi_1+\theta_1^2) &\quad (7)\\
\gamma(1) &= \phi_1\gamma(0) - \sigma^2\theta_1 &\quad (8)\\
\gamma(k) &= \phi_1\gamma(k-1)\,, \quad \forall k \geq 2 &\quad (9)
\end{aligned}$$

Substituting (8) into (7) gives:

$$\begin{aligned}
\gamma(0) &= \phi_1^2\gamma(0) - \sigma^2\theta_1\phi_1 + \sigma^2(1-\theta_1\phi_1+\theta_1^2)\\
\Rightarrow \gamma(0) &= \sigma^2\frac{1-2\theta_1\phi_1+\phi_1^2}{1-\phi_1^2}\\
\Rightarrow \gamma(1) &= \sigma^2\frac{(1-\theta_1\phi_1)(\phi_1-\theta_1)}{1-\phi_1^2}\,, \quad \text{[sub. } \gamma(0) \text{ into (8)]}
\end{aligned}$$

which, together with (9), implies

$$\gamma(k) = \sigma^2\phi^{k-1}\frac{(1-\theta_1\phi_1)(\phi_1-\theta_1)}{1-\phi_1^2}\,, \quad k \geq 1$$

Now, divide

$$\gamma(k) = \sigma^2 \phi^{k-1} \frac{(1-\theta_1\phi_1)(\phi_1-\theta_1)}{1-\phi_1^2}\,, \quad k \geq 1$$

by $\gamma(0)$ to get:

$$\rho(k) = \frac{\cancel{\sigma^2}\phi^{k-1}(1-\theta_1\phi_1)(\phi_1-\theta_1)/\cancel{(1-\phi_1^2)}}{\cancel{\sigma^2}(1-2\theta_1\phi_1+\phi_1^2)/\cancel{(1-\phi_1^2)}}\,, \quad |\phi_1| < 1$$

i.e., for the ARMA(1,1) model, $\rho(k) = A\phi^{k-1}$ behaves similar to $AR(1)$ (decays exponentially).

The Box-Jenkins methodology for forecasting

1. **Model identification**
    - Look at data. Compute sample ACF. Try to deduce whether model is $AR(p), MA(q), ARMA(p, q)$; decide on reasonable values for p and q.
2. **Parameter estimation**
3. **Verification**

    Check model obtained from 1 & 2

    - Good? Goto 4
    - Bad? Goto 1 & decide on new model
4. **Forecasting**