FORECASTING
STAT0010

Alexandros Beskos

a.beskos@ucl.ac.uk

'Lecture 4' Outline

1. Identification (so far)
2. Backshift operator
3. Invertibility
4. Box-Jenkins
5. Invertibility
6. Non-stationary processes

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 sine wave
$MA(1)$	$\rho(1) = \frac{-\theta_1}{1+\theta_1^2}$: '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 $\geq q + 1$
$ARMA(p, q)$	exponential decay or damped sine wave

(We'll see later how to distinguish between AR and $ARMA$.)

Remark 1

(For our purposes) It can be shown that for a large sample size T*:*

$$r(k) \overset{approx}{\sim} \mathcal{N}(0, 1/T)\,,$$

for any k *such that the theoretical ACF* $\rho(k) = 0$*. This means that we can use the* 95% *confidence interval* $(-1.96/\sqrt{T}, 1.96/\sqrt{T})$ *(c.f. white noise remark, lecture 1) to test whether a given ACF value is a 'genuine' zero or not. I.e.,*

- *for each* k *s.t.* $\rho(k) = 0$*, we expect* 95% *of realisations of this time series to have* $r(k)$ *inside interval*
- *observed values of* $r(k)$ *that fall outside these limits are considered 'significantly' different from zero at the 5% level*
- *expect to get* 5% *of* $r(k)$ *coeffs (for* k *s.t.* $\rho(k) = 0$*) outside the* 95% *confidence limits*

Definition 2 (Backshift operator)

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

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

By convention, $B := B^1$. Note that, e.g.

$$B^2 Y_t = BBY_t = B(BY_t) = B(Y_{t-1}) = BY_{t-1} = Y_{t-2}\,.$$

Example 3

In backshift notation, the $AR(p)$ model $Y_t = \epsilon_t + \sum_{j=1}^{p} \phi_j Y_{t-j}$ can be written as $\phi(B) Y_t = \epsilon_t$, where $\phi(B)$ is the $AR(p)$ characteristic polynomial.

To see this, write $AR(p)$ as $Y_t = \epsilon_t + \sum_{j=1}^{p} \phi_j B^j Y_t$. I.e.

$$\begin{aligned} Y_t - \sum_{j=1}^{p} \phi_j B^j Y_t &= \epsilon_t \\ \left(1 - \sum_{j=1}^{p} \phi_j B^j\right) Y_t &= \epsilon_t \\ \phi(B) Y_t = \epsilon_t \end{aligned}$$

Similarly (c.f. notes, Section 4.7), the $MA(q)$ model can be written as $Y_t = \theta(B)\epsilon_t$, where $\theta(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$ is the characteristic polynomial of the $MA(q)$ model.

Remark 4

The $ARMA(p,q)$ model can be written as $\phi(B)Y_t = \theta(B)\epsilon_t$.

Proposition 5

Let B be the backshift operator. Let $a, c \in \mathbb{R}$, and $\{X_t\}$ be some sequence. Then, $\forall i, j \in \mathbb{N} = \{0, 1, 2, \ldots\}$:

1. $Bc = c$
2. $B^j(cX_t) = cB^jX_t = cX_{t-j}$
3. $(aB^i + cB^j)X_t = aB^iX_t + cB^jX_t = aX_{t-i} + cX_{t-j}$
4. $B^iB^jX_t = B^iX_{t-j} = X_{t-j-i} = B^{i+j}X_t$
5. $(1 - aB)^{-1}X_t = \sum_{j=0}^{\infty} a^jB^jX_t = \sum_{j=0}^{\infty} a^jX_{t-j}\,, \quad \text{if } |a| < 1$

Proof 1 – 4: by definition(!). For 5 want: $(1-aB)^{-1}X_t = \sum_{j=0}^{\infty} a^j B^j X_t$.
'Proof' Consider:

$$
\begin{aligned}
(1-aB)\left(\sum_{j=0}^{\infty} a^j B^j\right) X_t &= \left(\sum_{j=0}^{\infty} a^j B^j - \sum_{j=0}^{\infty} a^{j+1} B^{j+1}\right) X_t \\
&= \left(\sum_{j=0}^{\infty} a^j B^j - \sum_{j=1}^{\infty} a^j B^j\right) X_t \\
&= \left(1 + \cancel{\sum_{j=1}^{\infty} a^j B^j} - \cancel{\sum_{j=1}^{\infty} a^j B^j}\right) X_t \\
&= X_t
\end{aligned}
$$

$\therefore \quad \sum_{j=0}^{\infty} a^j B^j = (1-aB)^{-1}$, provided sum converges, i.e. $|a| < 1$ ■

Remark 6

It can be shown that other expressions involving the backshift operator can often be manipulated as if B were a number or variable...

Example 7

An $AR(2)$ model can be represented as an $MA(\infty)$ model.

$AR(2)$: $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$ can be written, in terms of the backshift operator, as:

$$(1 - \phi_1 B - \phi_2 B^2) Y_t = \epsilon_t \, .$$

Write the $AR(2)$ characteristic polynomial as

$$\phi(B) = (1 - \lambda_1 B)(1 - \lambda_2 B) \, ,$$

—here λ_1^{-1} and λ_2^{-1} are roots of $\phi(B)$. Then

$$\begin{aligned} (1 - \lambda_1 B)(1 - \lambda_2 B) Y_t &= \epsilon_t \\ Y_t &= \left((1 - \lambda_1 B)(1 - \lambda_2 B)\right)^{-1} \epsilon_t \\ Y_t &= \frac{1}{(1 - \lambda_1 B)(1 - \lambda_2 B)} \epsilon_t \, , \end{aligned}$$

Now use partial fractions to deal with the RHS.

Partial fractions: find α_1, α_2, s.t.

$$\frac{1}{(1-\lambda_1 x)(1-\lambda_2 x)} = \frac{\alpha_1}{1-\lambda_1 x} + \frac{\alpha_2}{1-\lambda_2 x}$$

Multiply both sides by denominator of LHS; then put $x = \lambda_1^{-1}$ to get $\alpha_1 = \lambda_1(\lambda_1 - \lambda_2)^{-1}$ and put $x = \lambda_2^{-1}$ to get $\alpha_2 = -\lambda_2(\lambda_1 - \lambda_2)^{-1}$ i.e.

$$\frac{1}{(1-\lambda_1 x)(1-\lambda_2 x)} = \frac{1}{\lambda_1 - \lambda_2}\left(\frac{\lambda_1}{1-\lambda_1 x} - \frac{\lambda_2}{1-\lambda_2 x}\right)$$

I.e. $$Y_t = \frac{1}{\lambda_1 - \lambda_2}\left(\lambda_1(1-\lambda_1 B)^{-1} - \lambda_2(1-\lambda_2 B)^{-1}\right)\epsilon_t$$

Now, provided $|\lambda_1|, |\lambda_2| < 1$:

$$\begin{aligned} Y_t &= \frac{1}{\lambda_1 - \lambda_2}\left(\lambda_1 \sum_{j=0}^{\infty} \lambda_1^j B^j - \lambda_2 \sum_{j=0}^{\infty} \lambda_2^j B^j\right)\epsilon_t \\ &= \frac{1}{\lambda_1 - \lambda_2}\sum_{j=0}^{\infty}(\lambda_1^{j+1} - \lambda_2^{j+1})B^j \epsilon_t \\ &= \frac{1}{\lambda_1 - \lambda_2}\sum_{j=0}^{\infty}\left(\lambda_1^{j+1} - \lambda_2^{j+1}\right)\epsilon_{t-j}\,. \qquad \blacksquare \end{aligned}$$

Theorem 8 (roots of $AR(p)$ char. eqn. outside unit circle $\Leftrightarrow$ stationarity)

Consider $AR(p)$ process: $Y_t = \epsilon_t + \sum_{j=1}^{p} \phi_j Y_{t-j}$. Then $\{Y_t\}$ is (weakly) stationary iff all roots of AR characteristic equation:

$$\phi(x) := 1 - \phi_1 x - \phi_2 x^2 - \ldots - \phi_p x^p \,,$$

are outside unit circle.

'Sketch proof'

Let $\phi(B) Y_t = \epsilon_t$, with roots $\lambda_1^{-1}, \ldots, \lambda_p^{-1}$, i.e.

$$\lambda_0 (1 - \lambda_1 B)(1 - \lambda_2 B) \cdots (1 - \lambda_p B) Y_t = \epsilon_t$$

Now $(1 - \lambda_n B)^{-1}$ exists $\Leftrightarrow \sum_{j=0}^{\infty} \lambda_n^j B^j$ converges $\Leftrightarrow$ roots $|\lambda_n^{-1}| > 1$.

In this case we are allowed to write

$$Y_t = \lambda_0^{-1} (1 - \lambda_1 B)^{-1} (1 - \lambda_2 B)^{-1} \cdots (1 - \lambda_p B)^{-1} \epsilon_t$$

A result from calculus gives that if $\sum a_j^2 < \infty$, $\sum b_j^2 < \infty$ then $\sum (a * b)_j^2 < \infty$, where $(a * b)_j = \sum_{k=0}^{j} a_k b_{j-k}$.

This is useful because we have

$$(1 - \lambda_{p-1} B)^{-1} (1 - \lambda_p B)^{-1} \epsilon_t = \sum_{j=0}^{\infty} (a * b)_j \epsilon_{t-j}$$

with $a_j = \lambda_p^j$ and $b_j = \lambda_{p-1}^j$.

Thus Wolds theorem gives that $(1 - \lambda_{p-1} B)^{-1} (1 - \lambda_p B)^{-1} \epsilon_t$ is stationary.

Working similarly we obtain that Y_t is a stationary process.

Remark 9

1. *for finite order $AR(p)$, stationarity imposes restrictions on the parameters $\phi_1, \ldots, \phi_p$.*
2. *for finite order $MA(q)$, stationarity* **does not** *impose any restrictions on $\theta_1, \ldots, \theta_q$. However, ...*

Example 10

Consider the two $MA(1)$ models

$$\begin{aligned} \textit{model } \mathrm{I}: \quad Y_t &= \epsilon_t - \theta_1 \epsilon_{t-1} \\ \textit{model } \mathrm{II}: \quad Y_t &= \epsilon_t - \theta_1^{-1} \epsilon_{t-1} \end{aligned}$$

Recall model I has ACF with $\rho_{\mathrm{I}}(1) = -\theta_1(1+\theta_1^2)^{-1}$ and $\rho_{\mathrm{I}}(k) = 0\,, \forall |k| \geq 2$. Model II has

$$\rho_{\mathrm{II}}(1) = \frac{-\theta_1^{-1}}{1+\theta_1^{-2}} = \frac{-\theta_1}{1+\theta_1^2} = \rho_{\mathrm{I}}(1)\,.$$

and $\rho_{\mathrm{II}}(k) = 0\,, \forall\, |k| \geq 2$. Hence $\rho_{\mathrm{I}} \equiv \rho_{\mathrm{II}}$. This is **not** a nice situation!

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**
   - Using the model and values of (the model orders) p and q from the first step, estimate the unknown parameters, $\phi_1, \phi_2, \ldots, \phi_p$, $\theta_1, \ldots, \theta_q$, etc. and the variance of the white noise.

3. **Verification**

   Check model obtained from 1 & 2

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

4. **Forecasting**

This uniqueness problem can be solved by asking the question: Under what conditions can we represent an *MA* model as an *AR* model?

Definition 11

An MA model is said to be invertible if it can be written as an AR model.

Theorem 12 (roots of $MA(q)$ char. eqn. outside unit circle $\Leftrightarrow$ invertible)

Consider $MA(q)$ process: $Y_t = \epsilon_t - \sum_{j=1}^{p} \theta_j \epsilon_{t-j}$. Then $\{Y_t\}$ is invertible iff all roots of MA characteristic equation:

$$\theta(x) := 1 - \theta_1 x - \theta_2 x^2 - \ldots - \theta_q x^q ,$$

are outside unit circle.

Proof See Theorem 8.

Example 13

$MA(1)$: $Y_t = \epsilon_t - \theta_1 \epsilon_{t-1} =: \theta(B)\epsilon_t$

$Y_t = (1-\theta_1 B)\epsilon_t$. I.e if the $MA(1)$ root $|\theta_1^{-1}| > 1$ then

$$\epsilon_t = (1-\theta_1 B)^{-1} Y_t = \sum_{j=0}^{\infty} (\theta_1 B)^j Y_t = \sum_{j=0}^{\infty} \theta_1^j Y_{t-j}$$

Recall the two $MA(1)$ models:

$$\begin{array}{lccll} \text{model I:} & Y_t & = & \epsilon_t - \theta_1 \epsilon_{t-1}\,, & \text{root } = \theta_1^{-1} \\ \text{model II:} & Y_t & = & \epsilon_t - \theta_1^{-1} \epsilon_{t-1}\,, & \text{root } = \theta_1 \end{array}$$

Then, for $|\theta_1^{-1}| > 1$, model I is invertible and model II is not. The invertibility condition ensures ACF of MA models is unique (this holds for general $MA(q)$ models, not just for $MA(1)$).

Theorem 14 (Summary of Theorems 8 & 12)

Let Y_t *satisfy an* $ARMA(p,q)$ *model, i.e.* $\phi(B)Y_t = \theta(B)\epsilon_t$.

- *If the roots of the* $AR(p)$ *characteristic polynomial* $\phi(B)$ *are outside unit circle then* Y_t *is stationary; and*

$$\exists\, \psi(B) = \frac{\theta(B)}{\phi(B)}, \quad s.t. \;\; Y_t = \psi(B)\epsilon_t$$

- *If the roots of the* $MA(q)$ *characteristic polynomial* $\theta(B)$ *are outside unit circle then* Y_t *is invertible; and*

$$\exists\, \pi(B) = \frac{\phi(B)}{\theta(B)}, \quad s.t. \;\; \epsilon_t = \pi(B)Y_t$$

Example 15

See exercise sheet 4, questions 1 and 2!

Example 16 (*MA* representation of $ARMA(1,1)$)

A stationary $ARMA(1,1)$ model can written as

$$Y_t = \left(1 + (\phi_1 - \theta_1)\sum_{j=1}^{\infty} \phi_1^{j-1} B^j\right) \epsilon_t$$

To see this, write $ARMA(1,1)$ as $(1-\phi_1 B)Y_t = (1-\theta_1 B)\epsilon_t$, i.e.

$$Y_t = (1-\theta_1 B)(1-\phi_1 B)^{-1}\epsilon_t\,.$$

Assume $|\phi_1| < 1$ and the rest is left as an exercise.

Remark 17

Note also, that writing the model in terms of $\psi(B)\epsilon_t$ offers an alternative way to compute the ACF — recall lecture 3 used Yule-Walker for $ARMA(1,1)$.

In e.g. above, write $Y_t = \sum_{j=0}^{\infty}\psi_j\epsilon_t$ with

$$\psi_j = \begin{cases} 1\,, & j = 0 \\ (\phi_1 - \theta_1)\phi_1^{j-1}\,, & \text{oth.} \end{cases}$$

and use (from Wold) $\gamma(k) = \sigma^2\sum_{j=0}^{\infty}\psi_j\psi_{j+k}$.

Dealing with non-stationarity

Example 18

Consider the process

$$Y_t = \alpha_0 + \alpha_1 t\,, \quad \alpha_0, \alpha_1 \in \mathbb{R}\,.$$

In this case Y_t *is not stationary, since (the 'mean function')*

$$\mathbb{E}Y_t = \alpha_0 + \alpha_1 t\,,$$

which depends on t.

But, we can 'transform' the original non-stationary process to a stationary one by differencing. Consider:

$$\begin{aligned} Z_t &:= Y_t - Y_{t-1} \\ &= \cancel{\alpha_0} + \cancel{\alpha_1 t} - (\cancel{\alpha_0} + \alpha_1(\cancel{t} - 1)) \\ &= \alpha_1\,, \end{aligned}$$

which is now stationary.

Example 19

Consider an ARMA model where one (and only one) of the AR roots is equal to 1 and all the other roots are outside the unit circle. I.e. AR polynomial takes the form $\phi^*(B) := \phi(B)(1-B)$, where all the roots of $\phi(B)$ are outside the unit circle. We have:

$$\phi^*(B)Y_t = \phi(B)(1-B)Y_t = \theta(B)\epsilon_t$$

Again, define $Z_t := Y_t - Y_{t-1}$. I.e. $Z_t = (1-B)Y_t$, and we have

$$\phi(B)Z_t = \theta(B)\epsilon_t\,.$$

and Z_t is stationary.

Lemma 20

Y_t stationary $\Rightarrow$ $Y_t - Y_{t-1}$ stationary.

Proof $Y_t = \dfrac{\theta(B)}{\phi(B)}\epsilon_t$. Put $Z_t := Y_t - Y_{t-1}$:

$$\begin{aligned} Z_t &= (1-B)\frac{\theta(B)}{\phi(B)}\epsilon_t \\ \phi(B)Z_t &= (1-B)\theta(B)\epsilon_t \end{aligned}$$

Definition 21

Let $\{Y_t\}$ be some process. Then the difference operator (a.k.a backward difference operator) ∇ is defined by

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

The dth difference operator ∇^d is defined as

$$\nabla^d Y_t = \nabla^{d-1}\nabla Y_t\,, \quad \textit{where } \nabla^1 := \nabla\,.$$

Example 22

$$\nabla^2 Y_t = \nabla\nabla Y_t = \nabla(Y_t - Y_{t-1}) = Y_t - 2Y_{t-1} + Y_{t-2}\,.$$

Remark 23

Note that the dth difference operator can be written in terms of the backshift operator:

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

Example 24

$$\nabla^2 Y_t = (1-B)^2 Y_t = (1 - 2B + B^2)Y_t = Y_t - 2Y_{t-1} + Y_{t-2}\,.$$

Example 25

Consider $AR(1)$ model

$$Y_t = \phi_1 Y_{t-1} + \epsilon_t\,, \quad \epsilon \sim \mathcal{WN}(0, \sigma^2)\,.$$

Stationary if $|\phi_1| < 1$. But, if $\phi_1 = 1$:

$$Y_t = Y_{t-1} + \epsilon_t\,, \quad \text{[random walk]}$$

which is non-stationary (exercise: prove using successive substitutions and using $Y_0 = 0$)

However, note the difference

$$Z_t := \nabla Y_t = \epsilon_t \sim \mathcal{WN}(0, \sigma^2)\,,$$

is stationary.

Example 26

If an $ARMA$ model has d-many AR roots $= 1$, then

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

Now define $Z_t := (1 - B)^d Y_t = \nabla^d Y_t$ (i.e. take dth difference of Y_t). Then Z_t is stationary.

Remark 27

- *In practice, 1st differencing is often found to be adequate to make a series stationary.*
- *Sometimes, 2nd differencing is required.*
- *3rd or higher differencing is not usually required.*

CAVEAT!

Over-differencing will cause non-invertibility!

Example 28

Consider (again) random walk $Y_t = Y_{t-1} + \epsilon_t$*. Then first difference* $\nabla Y_t = \epsilon_t$ *is (stationary and) invertible. The second difference* $\nabla^2 Y_t = \nabla \epsilon_t = \epsilon_t - \epsilon_{t-1}$ *(is still stationary — c.f. Lemma 20) but is now not invertible — it is* $MA(1)$ *with* $\theta_1 = 1$*.*