FORECASTING
STAT0010

Alexandros Beskos

a.beskos@ucl.ac.uk

'Lecture 2' Outline

1. Stationarity
2. Moving average processes
3. Autoregressive processes

- Many forecasters are persuaded of the benefits of **parsimony**: *using as few parameters as possible!*
- Although complicated models can track data very well over the historical period for which parameters are estimated, they often perform poorly when used for out-of-sample forecasting!

The Box-Jenkins methodology for forecasting

1. **Model identification**
2. **Parameter estimation**
3. **Verification**

   Check model obtained from 1 & 2

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

4. **Forecasting**

Definition 1

$\{Y_t\}$ is weakly stationary if $\forall k, s, t$:

1. $\mathbb{E}(Y_s) = \mathbb{E}(Y_t)$, *i.e.* $\mu(t) = \mu$ *(a constant)*
2. $\text{cov}(Y_t, Y_{t+k}) = \text{cov}(Y_s, Y_{s+k})$

Remark 2

2 $\Rightarrow \gamma(t, t+k) = \gamma(s, s+k)$ *is a function of lag* k *and we write (define)*

$$\gamma(k) := \text{cov}(Y_t, Y_{t+k})$$

In particular, note that 2 $\Rightarrow \text{var}(Y_t) = \gamma(0)$, *i.e.* $\sigma_Y^2(t) = \sigma_Y^2$ *(const.)*

Remark 3 (...in other words...)

For stationary time series, the autocovariance

$$\text{cov}(Y_t, Y_s) =: \gamma(t, s) =: \gamma(s-t)$$

is a function of lag k*. I.e.*

$$\gamma(k) = \text{cov}(Y_t, Y_{t+k})$$

Recall $\{Y_t\}$ is strictly stationary if $\forall k, m, t_1, \ldots t_m$:

$$(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_m}) \stackrel{d}{=} (Y_{t_1+k}, Y_{t_2+k}, \ldots, Y_{t_m+k})$$

Caveat: 'Second order stationarity'

Notes says weak stationarity is also called second-order stationarity (c.f. notes, Defn. 3.2) But, **this is ambiguous!**.

Definition 4 ('Second order stationary')

$\{Y_t\}$ is sometimes called second order stationary or stationary to the 2nd order if $\forall k, s, t$,

$$(Y_t) \stackrel{d}{=} (Y_{t+k}) \quad \textit{and} \quad (Y_t, Y_s) \stackrel{d}{=} (Y_{t+k}, Y_{s+k})$$

Definition 5 ('Second order stationary')

Sometimes second order stationary = weakly stationary. (!)

We will use the 'weakly stationary' definition (Defn. 1).

Example 6

A non-stationary process (simplest) is given by

$$Y_t = \underbrace{\beta t}_{deterministic\ \ component} + \underbrace{\epsilon_t}_{stochastic\ \ component}, \quad \epsilon_t \sim WN$$

- $\mathbb{E}[Y_t] = \beta t$ *depends on* t
- *However,* $X_t = Y_t - \beta t$ *is weakly stationary.*

Example 7

Random Walk example

$$Y_t = Y_{t-1} + \epsilon_t, \;\; \epsilon_t \sim WN, \;\; Y_0 \;\; constant$$

1. *Solving recursively:*

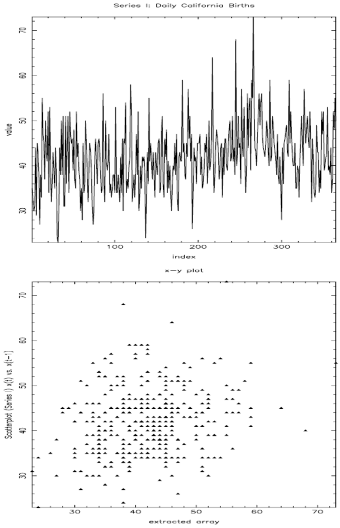
$$Y_t = \sum_{j=1}^{t} \epsilon_j + Y_0$$

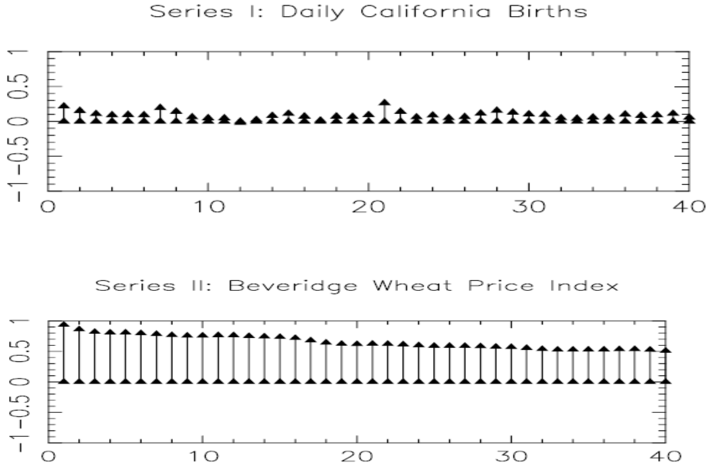
2. $\mathbb{E}\left[Y_t\right] = Y_0$ *time-invariant mean.*
3. $\mathbb{V}\mathrm{ar}\left[Y_t\right] = t\sigma^2$ *time dependent*
4. $X_t = Y_t - Y_{t-1}$ *is weakly stationary*

Fundamentals of the Sample Correlogram:
Time series often can exhibit correlation over time (sometimes known as serial correlation or autocorrelation). In the next figure we see a scater plot of $x(t)$ vs. $x(t-1)$, i.e. x with itself lagged by one time unit, for two time series.

- The one on the left for “Daily Californian Births” demonstrates weak linear predictability; and
- The series on the right for “log Wheat Prices Index” demonstrates strong linear predictability.

We can learn qualitative and quantitative information regarding linear predictability of a time series by looking at a correlogram (as demonstrated next)!


Series I: Daily California Births
value
index
Series II: Log Wheat Price Index
value
index
x–y plot
Scatterplot (Series I) x(t) vs. x(t−1)
extracted array
x–y plot
Scatterplot (Series II) x(t) vs. x(t−1)
extracted array


Series I: Daily California Births
−1 −0.5 0 0.5 1
0 10 20 30 40
Series II: Beveridge Wheat Price Index
−1 −0.5 0 0.5 1
0 10 20 30 40

Theorem 8 (Wold decomposition)

A weakly stationary time series $\{Y_t\}$ *can be represented as:*

$$Y_t = Z_t + \sum_{j=0}^{\infty} \beta_j \epsilon_{t-j}\,, \qquad \textit{where} \qquad \sum_{j=0}^{\infty} \beta_j^2 < \infty\,, \qquad \textit{with } \beta_0 = 1$$

and $\{Z_t\}$ *is a deterministic process.*

Proof Omitted. In this course: concentrate on the case $Z_t = \mu \in \mathbb{R}$.

Corollary 9

1. $E(Y_t) = \mu$
2. $\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \beta_j \beta_{j+k}$
3. $\rho(k) = \dfrac{\sum_{j=0}^{\infty} \beta_j \beta_{j+k}}{\sum_{j=0}^{\infty} \beta_j^2}$

Proof For (1):

$$Y_t = \sum_{j=0}^{\infty} \beta_j \epsilon_{t-j}, \qquad \Rightarrow \qquad \mathbb{E}(Y_t) = \sum_{j=0}^{\infty} \beta_j \underbrace{\mathbb{E}(\epsilon_{t-j})}_{0} = 0 \Rightarrow \text{①}$$

For (2):

$$\begin{aligned} \gamma(k) &= \text{cov}\left(\sum_{j=0}^{\infty}\beta_j\epsilon_{t-j}, \sum_{i=0}^{\infty}\beta_i\epsilon_{t+k-i}\right) \\ &= \sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\beta_j\beta_i \, \text{cov}(\epsilon_{t-j}, \epsilon_{t+k-i}) \qquad \text{(from cov property 7)} \end{aligned}$$

Now, $\{\epsilon_t\}$ has uncorrelated terms, i.e., recall:

$$\text{cov}(\epsilon_t, \epsilon_{t+k}) = \delta_{0,k}\sigma^2$$

Hence,

$$\text{cov}(\epsilon_{t-j}, \epsilon_{t+k-i}) = \delta_{t-j,t+k-i}\sigma^2 = \delta_{-j,k-i}\sigma^2 = \delta_{i,j+k}\sigma^2$$

$$\therefore \gamma(k) = \sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\beta_j\beta_i\sigma^2\delta_{i,j+k} = \sigma^2\sum_{j=0}^{\infty}\beta_j\beta_{j+k} \Rightarrow \text{②}$$

For (3): $\rho(k) = \gamma(k)/\gamma(0)$. I.e.

$$\gamma(0) = \gamma(k)|_{k=0} = \left(\sigma^2 \sum_{j=0}^{\infty} \beta_j \beta_{j+k}\right)\Bigg|_{k=0}$$

$$= \sigma^2 \sum_{j=0}^{\infty} \beta_j^2$$

$$\therefore \rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\cancel{\sigma^2} \sum_{j=0}^{\infty} \beta_j \beta_{j+k}}{\cancel{\sigma^2} \sum_{j=0}^{\infty} \beta_j^2} \Rightarrow 3 \qquad \blacksquare$$

Remark 10

Given a (non-stationary) time series $\{\tilde{Y}_t\}$ with non-zero constant mean μ, simply consider $\{Y_t\} = \{\tilde{Y}_t - \mu\}$ with zero mean. (For much of what follows, assume $\mathbb{E}(Y_t) = 0$).

From Wold we had

$$Y_t = \epsilon_t + \sum_{j=1}^{\infty} \beta_j \epsilon_{t-j}$$

Now, for finitely many non-zero weights $\{\beta_j\}_1^q$, i.e. for $\beta_j = 0$, for $j > q$:

Definition 11 (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

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

An MA model defines a process which, at time t, involves a random event at time t plus weighted random events from near past. E.g. economic indicators.

Example 12 ($MA(1)$)

$$Y_t = \epsilon_t - \theta_1 \epsilon_{t-1}$$

It is instructive to derive the autocorrelation function ρ of an $MA(1)$ process. This can be done in (at least) three 'different' ways:

- *via Wold*
- *c.f. notes*
- *multiplying through by Y_{t-k} and taking $\mathbb{E}(\cdot)$*

Via Wold Write $MA(1)$ in similar form as Wold representation:

$$Y_t = \epsilon_t - \theta_1 \epsilon_{t-1} = \sum_{j=0}^{\infty} \beta_j \epsilon_{t-j} \quad \text{with } \beta_j := \begin{cases} 1, & j = 0 \\ -\theta_1, & j = 1 \\ 0, & j \geq 2 \end{cases}$$

From Wold, $\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \beta_j \beta_{j+k}$. Note, for $MA(1)$ that $\beta_{j+k} = 0$ for $j + k \geq 2$. I.e.

$$\gamma(k) = \sigma^2 \sum_{j=0}^{1-k} \beta_j \beta_{j+k}$$

$$
\begin{aligned}
\gamma(k) &= \sigma^2 \sum_{j=0}^{1-k} \beta_j \beta_{j+k} \\
&= \sigma^2 \left(\beta_0 \beta_k + \sum_{j=1}^{1-k} \beta_j \beta_{j+k} \right) \\
&= \sigma^2 \left(\beta_k + \sum_{j=1}^{1-k} \theta_j \theta_{j+k} \right) \\
&= \begin{cases} \sigma^2(1+\theta_1^2), & k = 0 \\ -\sigma^2 \theta_1, & k = 1 \\ 0, & k \geq 2 \end{cases} \\
[\gamma \text{ is symmetric}] &= \begin{cases} \sigma^2(1+\theta_1^2), & k = 0 \\ -\sigma^2 \theta_1, & |k| = 1 \\ 0, & |k| \geq 2 \end{cases}
\end{aligned}
$$

cf notes $\{Y_t\}$ is $MA(1)$, i.e.

$$\begin{aligned} Y_t &= \epsilon_t - \theta_1 \epsilon_{t-1} \\ Y_{t-k} &= \epsilon_{t-k} - \theta_1 \epsilon_{t-k-1} \end{aligned}$$

Hence

$$\begin{aligned} \mathrm{cov}(Y_t, Y_{t-k}) &= \mathrm{cov}(\epsilon_t - \theta_1 \epsilon_{t-1}, \epsilon_{t-k} - \theta_1 \epsilon_{t-k-1}) \\ &= \mathrm{cov}(\epsilon_t, \epsilon_{t-k}) - \theta_1 \,\mathrm{cov}(\epsilon_t, \epsilon_{t-k-1}) \\ &\quad -\theta_1 \,\mathrm{cov}(\epsilon_{t-1}, \epsilon_{t-k}) + \theta_1^2 \,\mathrm{cov}(\epsilon_{t-1}, \epsilon_{t-k-1}) \\ [\text{note } \mathrm{cov}(\epsilon_s, \epsilon_t) = \sigma^2 \delta_{s,t}] &= \sigma^2(\delta_{0,k} - \theta_1 \delta_{-1,k} - \theta_1 \delta_{1,k} + \theta_1^2 \delta_{0,k}) \\ &= \sigma^2(1 + \theta_1^2)\delta_{0,k} - \theta_1 \sigma^2 (\delta_{-1,k} + \delta_{1,k}) \\ &= \sigma^2(1 + \theta_1^2)\delta_{0,k} - \theta_1 \sigma^2 \delta_{1,|k|} \\ &= \begin{cases} \sigma^2(1 + \theta_1^2), & k = 0 \\ -\theta_1 \sigma^2, & |k| = 1 \\ 0, & \text{oth. } (|k| \geq 2) \end{cases} \end{aligned}$$

Example 13 ($MA(2)$)

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

Find ρ via Wold

Via Wold Write $MA(2)$ in similar form as Wold representation:

$$Y_t = \sum_{j=0}^{\infty} \beta_j \epsilon_{t-j} \quad \text{with } \beta_j := \begin{cases} 1, & j = 0 \\ -\theta_j, & j = 1, 2 \\ 0, & j \geq 3 \end{cases}$$

From Wold, $\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \beta_j \beta_{j+k}$. Note, for $MA(2)$ that $\beta_{j+k} = 0$ for $j + k \geq 3$. I.e.

$$\gamma(k) = \sigma^2 \sum_{j=0}^{2-k} \beta_j \beta_{j+k}$$

$$
\begin{aligned}
\gamma(k) &= \sigma^2 \sum_{j=0}^{2-k} \beta_j \beta_{j+k} \\
&= \sigma^2 \left(\beta_0 \beta_k + \sum_{j=1}^{2-k} \beta_j \beta_{j+k} \right) \\
&= \sigma^2 \left(\beta_k + \sum_{j=1}^{2-k} \theta_j \theta_{j+k} \right) \\
&= \begin{cases} \sigma^2(1 + \theta_1^2 + \theta_2^2), & k = 0 \\ \sigma^2(-\theta_1 + \theta_1 \theta_2), & k = 1 \\ -\sigma^2 \theta_2, & k = 2 \\ 0, & k \geq 3 \end{cases}
\end{aligned}
$$

$$\gamma(k) = \begin{cases} \sigma^2(1+\theta_1^2+\theta_2^2), & k=0 \\ \sigma^2(-\theta_1+\theta_1\theta_2), & |k|=1 \\ -\sigma^2\theta_2, & |k|=2 \\ 0, & |k|\geq 3 \end{cases}$$

$$\therefore \rho(k) = \frac{\gamma(k)}{\gamma(0)} = \begin{cases} 1, & |k|=0 \\ \dfrac{-\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2}, & |k|=1 \\ \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2}, & |k|=2 \\ 0, & |k|\geq 3 \end{cases}$$

Example 14 ($MA(q)$)

An $MA(q)$ process can be written $Y_t = \epsilon_t - \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$,

$$\text{c.f. Wold:} \qquad Y_t = \sum_{j=0}^{\infty} \beta_j \epsilon_{t-j}\,, \qquad \text{with } \beta_j = \begin{cases} 1, & j = 0 \\ -\theta_j & j = 1, 2, \ldots, q \\ 0, & \text{oth.} \end{cases}$$

$$\Rightarrow \mathbb{E}(Y_t) = 0\,, \quad \gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \beta_j \beta_{j+k}$$

Now, $\beta_j = 0\,, \forall j \geq q+1$, i.e. $\beta_{j+k} = 0\,, \forall j \geq q-k+1$

$$\begin{aligned} \gamma(k) &= \sigma^2 \sum_{j=0}^{q-k} \beta_j \beta_{j+k} \\ &= \sigma^2 \left(\beta_k + \sum_{j=1}^{q-k} \theta_j \theta_{j+k} \right) \end{aligned}$$

$$\gamma(k) = \sigma^2 \left(\beta_k + \sum_{j=1}^{q-k} \theta_j \theta_{j+k} \right)$$

In particular: $\gamma(0) = \sigma^2 \left(1 + \sum_{j=1}^{q} \theta_j^2\right)$. I.e.

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \begin{cases} 1, & k = 0 \\ \dfrac{-\theta_k + \sum_{j=1}^{q-k} \theta_j \theta_{j+k}}{1 + \sum_{j=1}^{q} \theta_j^2}, & k = 1, \ldots, q \\ 0, & k \geq q+1 \end{cases}$$

c.f. Eqn. (8) notes.

Remark 15 (ACF 'cut-off')

Note that the ACF of an MA(q) process 'cuts off' at $|k| = q + 1$ (is zero for $|k| > q$). Hence, ACF can be used as a model identification tool!

Definition 16 (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

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

An AR model defines a process which, at time t, depends linearly on past values: $Y_{t-1}, Y_{t-2}, \ldots$ (together with a random term).

Example 17 (ρ of $AR(1)$)

$$\begin{aligned} Y_t &= \phi_1 Y_{t-1} + \epsilon_t && (1) \\ Y_{t-1} &= \phi_1 Y_{t-2} + \epsilon_{t-1} && (2) \\ Y_{t-2} &= \phi_1 Y_{t-3} + \epsilon_{t-2} && (3) \end{aligned}$$

Now substitute (2) into (1):

$$\begin{aligned} Y_t &= \phi_1(\phi_1 Y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \phi_1^2 Y_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t && (4) \end{aligned}$$

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

Now substitute (3) into (24):

$$\begin{aligned} Y_t &= \phi_1^2(\phi_1 Y_{t-3} + \epsilon_{t-2}) + \phi_1 \epsilon_{t-1} + \epsilon_t \\ &= \phi_1^3 Y_{t-3} + \phi_1^2 \epsilon_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t \end{aligned}$$

By similar successive substitutions:

$$\begin{aligned} Y_t &= \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \cdots + \cancelto{0}{\lim_{j \to \infty} \phi_1^j Y_{t-j}} \\ &= \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}\,, \end{aligned}$$

if sum exists...

We have:

$$\begin{aligned}
\mathbb{E}(Y_t^2) &= \mathbb{E}\Big(\sum_{j=0}^{\infty}\phi_1^j\epsilon_{t-j}\Big)^2 \\
&= \mathbb{E}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\phi_1^j\phi_1^k\epsilon_{t-j}\epsilon_{t-k} \\
&= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\phi_1^j\phi_1^k\mathbb{E}\left(\epsilon_{t-j}\epsilon_{t-k}\right) \\
&= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\phi_1^j\phi_1^k\sigma^2\delta_{j,k} \\
&= \sigma^2\sum_{j=0}^{\infty}\phi_1^{2j}
\end{aligned}$$

i.e. sum converges (in the mean square, w.p. 1), if $|\phi_1| < 1$ (c.f. geometric progression). So, provided that $|\phi_1| < 1$, then $\{Y_t\}$ can be written in form $\sum_{j=0}^{\infty}\phi_1^j\epsilon_{t-j}$.

Remark 18 ($AR(1)$ is $MA(\infty)$!)

To summarise, $AR(1)$ can be written as:

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

This agrees with Wold decomposition under the condition:

$$\beta_j = \phi_1^j\,.$$

Hence, we have that an $AR(1)$ process can be written as an infinite order moving average process $MA(\infty)$.

Corollary 19 (mean and γ for $AR(1)$)

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

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

Proof from above remark and Wold: (PTO)

From Wold:

$$
\begin{aligned}
\gamma(k) &= \sigma^2 \sum_{j=0}^{\infty} \beta_j \beta_{j+k} \\
&= \sigma^2 \sum_{j=0}^{\infty} \phi_1^j \phi_1^{j+k}, \qquad [k \in \mathbb{N}, \text{ i.e. } k = 0, 1, 2, \ldots] \\
&= \sigma^2 \phi_1^k \sum_{j=0}^{\infty} (\phi_1^2)^j \\
&= \frac{\sigma^2 \phi_1^k}{1 - \phi_1^2} \qquad [\text{geo. prog., } |\phi_1| < 1] \quad \Rightarrow \text{②}
\end{aligned}
$$

In particular $\mathrm{var}(Y_t) = \gamma(0) = \sigma^2/(1 - \phi_1^2)$. Hence:

$$\rho(k) = \gamma(k)/\gamma(0) = \phi_1^k, \quad k \in \mathbb{N}$$

and ACF decays exponentially with increasing lag k.