FORECASTING
STAT0010

Alexandros Beskos

a.beskos@ucl.ac.uk

'Lecture 9' Outline

1. Forecasting ARMA
   - example: $ARMA(1, 1)$
2. Forecasting non-stationary (ARIMA) models
   - example: $ARIMA(0, 1, 1)$
   - example: $ARIMA(1, 1, 0)$
3. Measuring forecast performance
   - measures of deviation
   - example: share dealing
   - logarithmic scoring rule

From last week...

Corollary 1

Any stationary ARMA process $\{Y_t\}$, can be written as

$$Y_t - \mu = \sum_{j=0}^{\infty} \psi_j \, \epsilon_{t-j} \, .$$

Corollary 2

Let $\{Y_t\}$ be an invertible ARMA process. Then,

$$\mathbb{E}(\epsilon_{T+j} | Y_{1:T}) = \begin{cases} \epsilon_{T+j} \, , & j \leq 0 \\ 0 \, , & j > 0 \end{cases}$$

Corollary 3

Let $\{Y_t\}$ be a stationary and invertible ARMA process. Then,

$$\widehat{Y}_{T+h} = \mu + \sum_{j=h}^{\infty} \psi_j \, \epsilon_{T+h-j} \, , \qquad e_T(h) = \sum_{j=0}^{h-1} \psi_j \, \epsilon_{T+h-j} \, .$$

Corollary 4

Let Y_t be a stationary and invertible $ARMA(p,q)$ process and let the coefficients $\{\psi_j\}$ be defined as above. Then the h-step ahead forecast error, with origin T, satisfies:

$$
\begin{aligned}
\mathbb{E}(e_T(h)) &= 0\,, \\
\mathrm{var}(e_T(h)) &= \sigma^2 \sum_{j=0}^{h-1} \psi_j^2\,.
\end{aligned}
$$

Example 5

Let $\{Y_t\}$ be a stationary, invertible $ARMA(1,1)$ process with mean $\mathbb{E}(Y_t) = \mu$. Find:

1. $\widehat{Y}_{T+h}$ *[h-step ahead forecast with origin T]*
2. $e_T(h)$ *[h-step ahead forecast error]*
3. $\mathbb{E}(e_T(h))$ *[mean h-step ahead forecast error]*
4. $\mathrm{var}(e_T(h))$ *[variance of h-step ahead forecast error]*

1. $\widehat{Y}_{T+h}$:

$$ARMA(1,1): \qquad Y_t = \mu + \phi_1(Y_{t-1} - \mu) + \epsilon_t - \theta_1 \epsilon_{t-1}\,.$$

I.e.

$$Y_{T+h} = \mu + \phi_1(Y_{T+h-1} - \mu) + \epsilon_{T+h} - \theta_1 \epsilon_{T+h-1}\,.$$

Forecast is:

$$\begin{aligned}
\widehat{Y}_{T+h} &= \mathbb{E}(Y_{T+h}|Y_{1:T}) \\
&= \mathbb{E}\big(\mu + \phi_1(Y_{T+h-1} - \mu) + \epsilon_{T+h} - \theta_1 \epsilon_{T+h-1}\big|Y_{1:T}\big) \\
&= \mu + \phi_1(\widehat{Y}_{T+h-1} - \mu) + \cancelto{0,\,[h>0]}{\mathbb{E}(\epsilon_{T+h}|Y_{1:T})} - \theta_1 \mathbb{E}(\epsilon_{T+h-1}|Y_{1:T})
\end{aligned}$$

Now, [from Corollary 2]

$$\mathbb{E}(\epsilon_{T+h-1}|Y_{1:T}) = \begin{cases} \epsilon_T\,, & h = 1 \\ 0\,, & h > 1 \end{cases}$$

Hence, (noting $\widehat{Y}_T = Y_T$) we have:

$$\widehat{Y}_{T+h} = \begin{cases} \mu + \phi_1(Y_T - \mu) - \theta_1 \epsilon_T\,, & h = 1 \\ \mu + \phi_1(\widehat{Y}_{T+h-1} - \mu)\,, & h > 1 \end{cases}$$

Hence, answer to part 1 is:

$$\widehat{Y}_{T+h} = \begin{cases} \mu + \phi_1(Y_T - \mu) - \theta_1 \epsilon_T\,, & h = 1 \\ \mu + \phi_1(\widehat{Y}_{T+h-1} - \mu)\,, & h > 1 \end{cases} \qquad \blacksquare$$

However, in this particular case, we can go further. Note

$$\begin{aligned} \widehat{Y}_{T+2} &= \mu + \phi_1(\widehat{Y}_{T+1} - \mu) \\ &= \mu + \phi_1(\cancel{\mu} + \phi_1(Y_T - \mu) - \theta_1 \epsilon_T - \cancel{\mu}) \\ &= \mu + \phi_1^2(Y_T - \mu) - \phi_1 \theta_1 \epsilon_T\,. \end{aligned}$$

Similarly:

$$\begin{aligned} \widehat{Y}_{T+3} &= \mu + \phi_1(\widehat{Y}_{T+2} - \mu) \\ &= \mu + \phi_1(\cancel{\mu} + \phi_1^2(Y_T - \mu) - \phi_1 \theta_1 \epsilon_T - \cancel{\mu}) \\ &= \mu + \phi_1^3(Y_T - \mu) - \phi_1^2 \theta_1 \epsilon_T\,. \end{aligned}$$

By induction:

1 $$\widehat{Y}_{T+h} = \mu + \phi_1^h(Y_T - \mu) - \phi_1^{h-1} \theta_1 \epsilon_T\,. \qquad \blacksquare$$

2. $e_T(h)$: In general, to find error, recall it is convenient to find MA representation. $\{Y_t\}$ is $ARMA(1,1)$; hence

$$
\begin{aligned}
(1-\phi_1 B)(Y_t-\mu) &= (1-\theta_1 B)\epsilon_t \\
Y_t-\mu &= (1-\theta_1 B)(1-\phi_1 B)^{-1}\epsilon_t \\
&= (1-\theta_1 B)\sum_{j=0}^{\infty}(\phi_1 B)^j\epsilon_t \qquad [\{Y_t\} \text{ stationary}] \\
&= (1-\theta_1 B)\sum_{j=0}^{\infty}\phi_1^j B^j\epsilon_t \\
&= \left(\sum_{j=0}^{\infty}\phi_1^j B^j-\theta_1\sum_{k=0}^{\infty}\phi_1^k B^{k+1}\right)\epsilon_t \\
&= \left(1+\sum_{j=1}^{\infty}\phi_1^j B^j-\theta_1\sum_{k=0}^{\infty}\phi_1^k B^{k+1}\right)\epsilon_t \\
&= \left(1+\sum_{j=1}^{\infty}\phi_1^j B^j-\theta_1\sum_{\ell=1}^{\infty}\phi_1^{\ell-1} B^{\ell}\right)\epsilon_t \qquad [\ell=k+1]
\end{aligned}
$$

$$
\begin{aligned}
Y_t - \mu &= \left(1 + \sum_{j=1}^{\infty} \phi_1^j B^j - \theta_1 \sum_{\ell=1}^{\infty} \phi_1^{\ell-1} B^{\ell}\right)\epsilon_t \\
&= \left(1 + \sum_{j=1}^{\infty} \left(\phi_1^j - \theta_1 \phi_1^{j-1}\right) B^j\right)\epsilon_t \\
&= \epsilon_t + \sum_{j=1}^{\infty} \left(\phi_1^j - \theta_1 \phi_1^{j-1}\right) \epsilon_{t-j} \\
&= \epsilon_t + \sum_{j=1}^{\infty} (\phi_1 - \theta_1)\phi_1^{j-1} \epsilon_{t-j} \\
&=: \sum_{j=0}^{\infty} \psi_j \, \epsilon_{t-j}\,,
\end{aligned}
$$

where

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

We want $e_T(h) = Y_{T+h} - \widehat{Y}_{T+h}$, with $Y_{T+h} = \mu + \sum_{j=0}^{\infty} \psi_j \, \epsilon_{T+h-j}$. Recall [from Corollary 3 or c.f. lecture 8], that $\widehat{Y}_{T+h} = \mathbb{E}(Y_{T+h} | Y_{1:T})$

$$
\begin{aligned}
&= \mathbb{E}\left(\mu + \sum_{j=0}^{\infty} \psi_j \, \epsilon_{T+h-1} | Y_{1:T}\right) \\
&= \mu + \sum_{j=0}^{\infty} \psi_j \, \mathbb{E}(\epsilon_{T+h-j} | Y_{1:T}) \\
&= \mu + \sum_{j=h}^{\infty} \psi_j \, \epsilon_{T+h-j} \,, \qquad \left[\mathbb{E}(\epsilon_{T+h-j} | Y_{1:T}) = \begin{cases} \epsilon_{T+h-j} \,, & j \geq h \\ 0 \,, & j < h \end{cases}\right]
\end{aligned}
$$

Hence, [from Corollary 3 or c.f. lecture 8]

$$
\begin{aligned}
e_T(h) &= \mu + \sum_{j=0}^{\infty} \psi_j \, \epsilon_{T+h-j} - \mu - \sum_{j=h}^{\infty} \psi_j \, \epsilon_{T+h-j} \\
&= \sum_{j=0}^{h-1} \psi_j \, \epsilon_{T+h-j}
\end{aligned}
$$

$$e_T(h) = \sum_{j=0}^{h-1} \psi_j \, \epsilon_{T+h-j}$$

In the $ARMA(1,1)$ case,

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

Hence, for $ARMA(1,1)$:

2 $e_T(h) = \epsilon_{T+h} + (\phi_1 - \theta_1) \sum_{j=1}^{h-1} \phi_1^{j-1} \epsilon_{T+h-j}\,.$ ■

Therefore

3 $\mathbb{E}(e_T(h)) = 0\,.$ ■

And [from Corollary 4]

$$\mathrm{var}(e_T(h)) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 = \sigma^2 \left(1 + (\phi_1 - \theta_1)^2 \sum_{j=1}^{h-1} \phi_1^{2j-2} \right)$$

(Sum is geometric series)

4 $\mathrm{var}(e_T(h)) = \sigma^2 \left(1 + (\phi_1 - \theta_1)^2 \frac{1 - \phi_1^{2h-2}}{1 - \phi_1^2} \right)\,.$ ■

ARMA Forecasting Summary

1. Write model equation for Y_{T+h}
2. Put hats on everything ('hat', $\hat{\cdot}$ means expectation conditional on $Y_{1:T}$)
3. Calculate $\widehat{Y}_{T+1}$ using $\widehat{Y}_{T+j} = Y_{T+j}$, for $j \leq 0$, and
$$\hat{\epsilon}_{T+j} := \mathbb{E}(\epsilon_{T+j} | Y_{1:T}) = \begin{cases} 0\,, & j > 0 \\ \epsilon_{T+j}\,, & j \leq 0 \end{cases}$$
4. Calculate $\widehat{Y}_{T+2}$ using above (together with the value of $\widehat{Y}_{T+1}$)
5. Similarly, calculate $\widehat{Y}_{T+h}$
6. To find error properties, find sequence of coefficients ψ_j in the following *MA* representation of the model:
$$Y_t - \mu = \sum_{j=0}^{\infty} \psi_j \, \epsilon_{t-j}\,,$$
and use $\mathbb{E}(e_T(h)) = 0$, and $\text{var}(e_T(h)) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$

For non-stationary models, an *MA* representation does not exist —recall, e.g. $(1-\phi_1 B)Y_t = \epsilon_t \Rightarrow Y_t = (1-\phi_1^j B)^{-1}\epsilon_t = \sum_{j=0}^{\infty}\phi_1^j \epsilon_{t-j}$ which only exists (converges) if $|\phi_1| < 1$.
However, recall

$$\mathrm{var}(e_T(h)) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 \,,$$

i.e., variance of h-step error (with finite h) only depends on first h-many values of ψ_j. Therefore, we can proceed in a similar way as before.

Remark 6 (Caveat: non-stationary error variance increases without bound)

Recall that, for stationary models, as $h \to \infty$:

$$\mathrm{var}(e_T(h)) \to \mathrm{var}(Y_t) \,.$$

However for nonstationary models $\mathrm{var}(e_T(h))$ diverges. (Because $\sum_{j=0}^{\infty}\psi_j^2 \not< \infty$ for nonstationary models.)

Example 7 (Forecasting $ARIMA(0,1,1)$ models)

Find the h-step ahead forecast and associated error. Show mean error is zero. Find error variance.

1. Model equation:

$$(1-B)Y_{T+h} = (1-\theta_1 B)\epsilon_{T+h} \,.$$

I.e.

$$Y_{T+h} = Y_{T+h-1} + \epsilon_{T+h} - \theta_1 \epsilon_{T+h-1} \,.$$

2. Put hats on everything ($\hat{\cdot}$ means expectation conditional on $Y_{1:T}$):

$$\widehat{Y}_{T+h} = \widehat{Y}_{T+h-1} + \hat{\epsilon}_{T+h} - \theta_1 \hat{\epsilon}_{T+h-1} \,.$$

3. Calculate $\widehat{Y}_{T+1}$. Note, that $\widehat{Y}_T = Y_T$, $\hat{\epsilon}_{T+1} = 0$ and $\hat{\epsilon}_T = \epsilon_T$. Hence

$$\widehat{Y}_{T+1} = Y_T - \theta_1 \epsilon_T \,.$$

4. Calculate $\widehat{Y}_{T+2}$. Note, that $\hat{\epsilon}_{T+2} = 0$ and $\hat{\epsilon}_{T+1} = 0$. Hence

$$\widehat{Y}_{T+2} = \widehat{Y}_{T+1}$$

5. Similarly, for $h > 2$

$$\widehat{Y}_{T+h} = \widehat{Y}_{T+h-1} = \ldots = \widehat{Y}_{T+1}$$

6. Find *MA* representation, i.e. ψ_j, s.t. $Y_t = \sum_{j=0}^{\infty} \psi_j \, \epsilon_{t-j}$:

$$(1-B)Y_t = (1-\theta_1 B)\epsilon_t \,,$$

i.e.

$$\begin{aligned}
Y_t &= (1-\theta_1 B)(1-B)^{-1}\epsilon_t = (1-\theta_1 B)\sum_{j=0}^{\infty} B^j \epsilon_t \\
&= \left(\sum_{j=0}^{\infty} B^j - \theta_1 \sum_{k=0}^{\infty} B^{k+1}\right)\epsilon_t \\
&= \left(1 + \sum_{j=1}^{\infty} B^j - \theta_1 \sum_{k=0}^{\infty} B^{k+1}\right)\epsilon_t \\
&= \left(1 + \sum_{j=1}^{\infty} B^j - \theta_1 \sum_{k=1}^{\infty} B^{k}\right)\epsilon_t \qquad [k+1 \hookleftarrow k] \\
&= \left(1 + \sum_{j=1}^{\infty} (1-\theta_1)\right)\epsilon_{t-j}
\end{aligned}$$

I.e. $\psi_0 = 1$ and $\psi_j = 1 - \theta_1$ for $j > 0$.

$\psi_0 = 1$ and $\psi_j = 1 - \theta_1$ for $j > 0$. We have

$$\mathbb{E}(e_T(h)) = 0\,,$$

and

$$\begin{aligned}
\mathrm{var}(e_T(h)) &= \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 \\
&= \sigma^2 \left(1 + \sum_{j=1}^{h-1} (1 - \theta_1)\right) \\
&= \sigma^2 \left(1 + (1 - \theta_1) \sum_{j=1}^{h-1} 1\right) \\
&= \sigma^2 \left(1 + (h-1)(1 - \theta_1)\right). \qquad \blacksquare
\end{aligned}$$

Note, $h \to \infty$, $\mathrm{var}(e_T(h)) \not< \infty$.

Example 8 (Forecasting $ARIMA(1,1,0)$ models)

Find the h-step ahead forecast and associated error. Show mean error is zero. Find error variance.

1. Model equation:

$$(1-\phi_1 B)(1-B)Y_{T+h} = \epsilon_{T+h}\,.$$

Expand brackets:

$$Y_{T+h} = (1+\phi_1)Y_{T+h-1} - \phi_1 Y_{T+h-2} + \epsilon_{T+1}\,.$$

2. Put hats on everything:

$$\widehat{Y}_{T+h} = (1+\phi_1)\widehat{Y}_{T+h-1} - \phi_1\widehat{Y}_{T+h-2} + \hat{\epsilon}_{T+1}\,.$$

3. Calculate $\widehat{Y}_{T+1}$. Note, that $\widehat{Y}_T = Y_T$, $\widehat{Y}_{T-1} = Y_{T-1}$ and $\hat{\epsilon}_{T+1} = 0$:

$$\widehat{Y}_{T+1} = (1+\phi_1)Y_T - \phi_1 Y_{T-1}\,.$$

4. Calculate $\widehat{Y}_{T+2}$. Note that $\widehat{Y}_{T+1} \neq Y_{T+1}$, $\widehat{Y}_T = Y_T$ and $\hat{\epsilon}_{T+2} = 0$:

$$\widehat{Y}_{T+2} = (1+\phi_1)\widehat{Y}_{T+1} - \phi_1 Y_T\,.$$

5. Similarly, for $h > 2$ (n.b. error in course notes)

$$\widehat{Y}_{T+h} = (1+\phi_1)\widehat{Y}_{T+h-1} - \phi_1\widehat{Y}_{T+h-2}\,.$$

6. Find *MA* representation, i.e. ψ_j, s.t. $Y_t = \sum_{j=0}^{\infty} \psi_j \, \epsilon_{t-j}$:

$$(1 - \phi_1 B)(1 - B) Y_{T+h} = \epsilon_{T+h} \,,$$

i.e.

$$\begin{aligned}
Y_t &= \frac{1}{(1-\phi_1 B)(1-B)} \epsilon_t \\
&= \frac{1}{1-\phi_1} \left(\frac{1}{1-B} - \frac{\phi_1}{1-\phi_1 B} \right) \epsilon_t \qquad \text{[partial fractions]} \\
&= \frac{1}{1-\phi_1} \left(\sum_{j=0}^{\infty} B^j - \phi_1 \sum_{j=0}^{\infty} \phi_1^j B^j \right) \epsilon_t \\
&= \frac{1}{1-\phi_1} \sum_{j=0}^{\infty} (1 - \phi_1^{j+1}) B^j \epsilon_t \\
&= \sum_{j=0}^{\infty} \frac{1 - \phi_1^{j+1}}{1 - \phi_1} \, \epsilon_{t-j} \,.
\end{aligned}$$

I.e. $\psi_j = (1 - \phi_1^{j+1})/(1 - \phi_1)$.

$$\psi_j = (1 - \phi_1^{j+1})/(1 - \phi_1)\,.$$

we have

$$\mathbb{E}(e_T(h)) = 0\,,$$

and

$$\begin{aligned} \mathrm{var}(e_T(h)) &= \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 \\ &= \sigma^2 \sum_{j=0}^{h-1} \left(\frac{1 - \phi_1^{j+1}}{1 - \phi_1} \right)^2. \qquad \blacksquare \end{aligned}$$

Note, $h \to \infty$, $\mathrm{var}(e_T(h)) \not< \infty$.

Question 9 (measuring forecast performance)

Given, at time T, we forecast values $\hat{y}_{T+1}, \hat{y}_{T+2}, \ldots, \hat{y}_{T+h}$, and by time $T+h$, we observe actual values $y_{T+1}, y_{T+2}, \ldots, y_{T+h}$. How can we assess how good our forecasts are?

Definition 10 (error measures)

- *Mean square error*
$$FMSE_h(T) := \frac{1}{h}\sum_{j=1}^{h}(y_{T+j} - \hat{y}_{T+j})^2 .$$
- *Mean absolute error*
$$FMAE_h(T) := \frac{1}{h}\sum_{j=1}^{h}\left|y_{T+j} - \hat{y}_{T+j}\right| .$$
- *Mean absolute percentage error*
$$FMAPE_h(T) := \frac{1}{h}\sum_{j=1}^{h}\left|\frac{y_{T+j} - \hat{y}_{T+j}}{y_{T+h}}\right| .$$

All these error measures quantify average deviation from forecasts. However, the do not take into account uncertainty (error variance).

Example 11 (importance of forecast uncertainty!)

Think about buying 1000 *shares @* £3.20*/share. Forecasts for next month:*

Forecaster A: predicts £3.30 **exactly**
Forecaster B: predicts $£3.40 \pm 0.20$ *(error std. dev.)*

I.e.

Forecast A: $\mathbb{P}(price < 3.20) = 0$
Forecast B: $\mathbb{P}(price < 3.20) = 0.159$ *[Assuming price* $\sim \mathcal{N}(340, 20^2)$ *]*

Now suppose (unfortunately) that next month true price = £3.18*. Forecaster A definitely looses us money! Forecaster B might have put us off buying any shares, i.e. loss = 0! But measures* $\Rightarrow$ *A is better?!?*

measure	*A*	*B*
$FMSE_1$	144	484
$FMAE_1$	12	22
$FMAPE_1$	0.038	0.069

Probabilistically, a forecast is a probability distribution of Y_{T+h}, i.e. we have asserted that Y_{T+h} has some density function $f_{Y_{T+h}|Y_{1:T}}$, conditioned on $Y_{1:T}$, with mean $\hat{y}_{T+h}$ and variance $\text{var}(e_T(h))$.

E.g., let $\epsilon_T \sim \mathcal{N}(0, \sigma^2)$. Then $\{Y_t\}$ are normally distributed and the probability forecast asserts

$$Y_{T+h}|Y_{1:T} \sim \mathcal{N}\left(\hat{y}_{T+h}, \text{var}(e_T(h))\right).$$

I.e., density function

$$f_{Y_{T+h}|Y_{1:T}}(y) = \left(2\pi\,\text{var}(e_T(h))\right)^{-1/2} \exp\left(-\frac{(y-\hat{y}_{T+h})^2}{2\,\text{var}(e_T(h))}\right),$$

quantifies the relative likelihood that the actual value is drawn from our probability forecast. Now take $-\log(\cdot)$ (small value $\Rightarrow$ good forecast):

$$\begin{aligned} -\log f_{Y_{T+h}|Y_{1:T}}(y) &= -\log\left(\left(2\pi\,\text{var}(e_T(h))\right)^{-1/2}\right) + \frac{(y-\hat{y}_{T+h})^2}{2\,\text{var}(e_T(h))} \\ &= \frac{1}{2}\log 2\pi + \frac{1}{2}\log\text{var}(e_T(h)) + \frac{(y-\hat{y}_{T+h})^2}{2\,\text{var}(e_T(h))} \end{aligned}$$

I.e., for an actual value $y = y_{T+h}$, this is

$$-\log f_{Y_{T+h}|Y_{1:T}}(y_{T+h}) = \text{const.} + \frac{1}{2}\log \operatorname{var}(e_T(h)) + \frac{(y_{T+h} - \hat{y}_{T+h})^2}{2\operatorname{var}(e_T(h))}$$

Summing these quantities (without const.) over all h-many forecasts...

Definition 12

Define logarithmic scoring rule over forecasts $\hat{y}_{T+1:T+h}$, as

$$\mathcal{L}_h := \frac{1}{2}\sum_{j=1}^{h} \log \operatorname{var}(e_T(j)) + \frac{(y_{T+j} - \hat{y}_{T+j})^2}{\operatorname{var}(e_T(j))} .$$

Remark 13

Note:

- *Small $\mathcal{L}_h \Leftrightarrow$ 'good' forecast*
- *1st term: $\log \operatorname{var}(e_T(j))$ large $\Leftrightarrow$ error variance large*
- *2nd term: $\frac{(y_{T+j} - \hat{y}_{T+j})^2}{\operatorname{var}(e_T(j))}$ large $\Leftrightarrow$ squared error $\gg$ error variance*

If we apply logarithmic scoring rule to the share dealing Example 11, we get

measure	A	B
$FMSE_1$	144	484
$FMAE_1$	12	22
$FMAPE_1$	0.038	0.069
$\mathcal{L}_1$	∞	**3.60**

I.e., simple measures of deviation imply (incorrectly) that Forecaster A is better, whereas logarithmic rule (correctly) identifies Forecaster B is better.

Remark 14

In practice, the forecast performance measures can be used as part of the validation procedure, as follows.

- *split data up into two sets:*
  - *'fitting' (or 'training') set [# data points typically = 80% of total data; e.g., for $T = 100$, choose $Y_1, \ldots, Y_{80}$, say]*
  - *validation (or 'hold-out') set [on remaining data $Y_{81}, \ldots, Y_{100}$]*
- *use Box-Jenkins approach only on the* **training** *set to choose a few good candidate models*
- *for each model:*
  - *generate forecasts*
  - *check forecasting performance by comparing forecasts with your hold-out set*