FORECASTING
STAT0010

Alexandros Beskos

a.beskos@ucl.ac.uk

'Lecture 8' Outline

1. Box Jenkins, so far
2. Forecasting (AR)
3. Forecasting (MA)
4. Forecasting (ARMA)

The Box-Jenkins methodology for forecasting

1. **Model identification**
   - **Plot** data, **ACF**, and **PACF**. **Difference** if required. Try to deduce model type.
   - 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 (via **least squares** or otherwise) the unknown parameters, $\mu, \phi_1, \ldots, \phi_p$, $\theta_1, \ldots, \theta_k$, $\Phi_1, \ldots$, $\Theta_1, \Theta_2, \ldots$, etc.
3. **Verification**
   Check model obtained from 1 & 2 by looking at **residuals**. Use **ACF**, **PACF** **t-tests**, R^2, **Box-Pierce**, **AIC**, etc.

   - Good? Goto 4 (or goto 2 and try to **overfit**, etc.)
   - Bad? Goto 1 & decide on new model
4. **Forecasting**

Check list

1. plot series [`ts.plot`]
2. plot ACF [`acf`], PACF [`pacf`]
3. Use `describe` to check mean (constant terms?) & variance (if, e.g. $\text{var}(\nabla Y) >> \text{var}(Y)$ you might have overdifferenced), symmetry of distribution (e.g. t-tests and especially AIC work 'better' on normal distribution)
4. Check non-stationarity. If non-stat., take $\nabla_{?}^{??}$ and goto 1
5. Make a shortlist of a few candidate models. Use `arima` command to estimate model (and produce residuals)
6. Check standard errors of estimated parameters to check whether you have overfitted (i.e. used a model with too many parameters)
7. check residuals (plot time series, ACF, PACF), R^2, Ljung-Box-Pierce. (ACF or PACF might suggest why your model is wrong.)
8. If all looks good, try overfitting (and goto 6)
9. Pick a few of the better models and compute AIC, AIC_C, and/or BIC.
10. Give one, and only one, answer! Don't just follow this checklist! You can also use (a little bit of) your own initiative!!!

Definition 1

Consider series $Y_1, \ldots, Y_T$ up to time T. Assuming correct model of the data is known, the h-step ahead forecast is defined as

$$\widehat{Y}_{T+h} := \mathbb{E}(Y_{T+h} | Y_T, Y_{T-1}, \ldots, Y_1)\,, \quad h \geq 1\,.$$

Here, T is called the (forecast) origin, and h the horizon or lead-time.

Lemma 2

$\widehat{Y}_{T+h}$ minimises the mean square error:

$$MSE(c) := \mathbb{E}((Y_{T+h} - c)^2 | Y_T, \ldots, Y_1)\,.$$

'Proof' Let $\widehat{Y}$ be the prediction of some random variable Y, with $\mu := \mathbb{E}(Y)$. Then, the mean square error is

$$\begin{aligned} MSE(\widehat{Y}) &= \mathbb{E}((Y - \widehat{Y})^2) \\ &= \mathbb{E}\big(\,((Y - \mu) + (\mu - \widehat{Y}))^2\big) \\ &= \mathbb{E}((Y - \mu)^2) + 2(\mu - \widehat{Y})\cancelto{0}{\mathbb{E}(Y - \mu)} + (\mu - \widehat{Y})^2 \\ &= \mathrm{var}(Y) + (\mu - \widehat{Y})^2\,. \end{aligned}$$

$$MSE(\widehat{Y}) = \text{var}(Y) + (\mu - \widehat{Y})^2 \,.$$

But, $(\mu - \widehat{Y})^2 \geq 0$, with equality iff $\widehat{Y} = \mu = \mathbb{E}(Y)$. Hence,

$$MSE(\mathbb{E}(Y)) = \min_{\widehat{Y}} MSE(\widehat{Y}) \,,$$

i.e. the mean of the variable (that we are trying to predict) minimises the mean square prediction error.

To finish proof, put $Y = Y_{T+h}$ and replace $\mathbb{E}(\cdot)$ with $\mathbb{E}(\cdot | Y_T, \ldots, Y_1)$. ■

Alternative 'proof'

$$MSE(\widehat{Y}) = \mathbb{E}((Y - \widehat{Y})^2) \,,$$

Differentiate:

$$\begin{aligned} \frac{d}{d\widehat{Y}} \mathbb{E}((Y - \widehat{Y})^2) &= \mathbb{E}\left(\frac{d}{d\widehat{Y}} (Y - \widehat{Y})^2\right) \\ &= -2\,\mathbb{E}(Y - \widehat{Y}) \qquad \text{[chain rule]} \end{aligned}$$

Now put this equal to 0, to get $\widehat{Y} = \mathbb{E}(Y)$ and note 2nd derivative is positive, hence $\mathbb{E}(Y)$ minimises MSE. ■

Consider stationary $AR(1)$ process with non-zero mean:

$$Y_t = \mu + \phi_1(Y_{t-1} - \mu) + \epsilon_t \,.$$

Example 3 (One-step ahead forecast)

$$\begin{aligned}
\widehat{Y}_{T+1} &= \mathbb{E}(Y_{T+1}|Y_T, \ldots, Y_1) \\
&= \mathbb{E}(\mu + \phi_1(Y_T - \mu) + \epsilon_{T+1}|Y_T, \ldots, Y_1) \\
&= \mu + \phi_1\mathbb{E}(Y_T|Y_T, \ldots, Y_1) - \phi_1\mu + \cancelto{0}{\mathbb{E}(\epsilon_{T+1}|Y_T, \ldots, Y_1)} \\
&= \mu + \phi_1(Y_T - \mu)\,.
\end{aligned}$$

Consider stationary $AR(1)$ process with non-zero mean:

$$Y_t = \mu + \phi_1(Y_{t-1} - \mu) + \epsilon_t \,.$$

Example 4 (Two-step ahead forecast)

$$\begin{aligned}
\widehat{Y}_{T+2} &= \mathbb{E}(Y_{T+2}|Y_T, \dots, Y_1) \\
&= \mathbb{E}(\mu + \phi_1(Y_{T+1} - \mu) + \epsilon_{T+2}|Y_T, \dots, Y_1) \\
&= \mu + \phi_1\mathbb{E}(Y_{T+1}|Y_T, \dots, Y_1) - \phi_1\mu + \cancelto{0}{\mathbb{E}(\epsilon_{T+2}|Y_T, \dots, Y_1)} \\
&= \mu + \phi_1(\widehat{Y}_{T+1} - \mu) \\
&= \mu + \phi_1(\mu + \phi_1(Y_T - \mu) - \mu) \qquad \textit{[from Example 3]} \\
&= \mu + \phi_1^2(Y_T - \mu) \,.
\end{aligned}$$

Remark 5

The one and two step ahead forecasts for a stationary $AR(1)$ process with non-zero mean:

$$Y_t = \mu + \phi_1(Y_{t-1} - \mu) + \epsilon_t \,,$$

are

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

Similarly, the h-step ahead forecast is

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

Note, since $\{Y_t\}$ stationary, we have $|\phi_1| < 1$. Hence,

$$\widehat{Y}_{T+h} \to \mu \,, \quad \text{as } h \to \infty \,.$$

I.e., to make predictions for an infinite time into the future, the best we can do is use the mean of the series to make long term predictions.

$\widehat{Y}_{T+h} \to \mu$ holds for all stationary *ARMA* models.

Definition 6 (Forecast error)

Define the h-step ahead forecast error, with origin T, by

$$e_T(h) = Y_{T+h} - \widehat{Y}_{T+h}\,.$$

Example 7 ($AR(1)$ one-step ahead forecast is unbiased)

For the $AR(1)$ case,

$$\begin{aligned} Y_{T+1} &= \mu + \phi_1(Y_T - \mu) + \epsilon_{T+1} && \textit{[from model defn.]} \\ \widehat{Y}_{T+1} &= \mu + \phi_1(Y_T - \mu) && \textit{[from Example 3]} \end{aligned}$$

Hence,

$$e_T(1) = Y_{T+1} - \widehat{Y}_{T+1} = \epsilon_{T+1}\,,$$

i.e. $\mathbb{E}(e_T(1)) = \mathbb{E}(\epsilon_{T+1}) = 0$. *Therefore, the forecast* $\widehat{Y}_{T+1}$ *is unbiased.*

Example 8 ($AR(1)$ forecast variance)

$$\text{var}(e_T(1)) = \text{var}(\epsilon_{T+1}) = \sigma^2\,.$$

Remark 9 (Interpretation of $\mathbb{E}(e_T(1)) = 0$ and $\text{var}(e_T(1)) = \sigma^2$)

If you simulate AR(1) lots of times and estimate 1-step ahead forecast for each realisation, then,

- *the expected forecast error will be zero*
- *the forecast error variance will be σ^2*

$AR(1)$ errors for longer term forecasts can be found by representing $AR(1)$ as $MA(\infty)$. Recall:

$$\begin{aligned} Y_t - \mu &= \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j} . && [AR(1) \text{ as an } MA(\infty) \text{ model}] \\ \widehat{Y}_{T+h} &= \mu + \phi_1^h (Y_T - \mu) && [\text{from Remark 5}] \end{aligned}$$

Now,

$$\begin{aligned} e_T(h) &= Y_{T+h} - \widehat{Y}_{T+h} \\ &= \cancel{\mu} + \sum_{j=0}^{\infty} \phi_1^j \epsilon_{T+h-j} - (\cancel{\mu} + \phi_1^h (Y_T - \mu)) \end{aligned}$$

$$
\begin{aligned}
e_T(h) &= \sum_{j=0}^{\infty} \phi_1^j \epsilon_{T+h-j} - (\phi_1^h (Y_T - \mu)) \\
&= \sum_{j=0}^{\infty} \phi_1^j \epsilon_{T+h-j} - \phi_1^h \sum_{\ell=0}^{\infty} \phi_1^\ell \epsilon_{T-\ell} && [\text{write } Y_T - \mu \text{ as } MA(\infty)] \\
&= \sum_{j=0}^{\infty} \phi_1^j \epsilon_{T+h-j} - \phi_1^h \sum_{k=h}^{\infty} \phi_1^{k-h} \epsilon_{T+h-k} && [\text{put } \ell = k - h] \\
&= \sum_{j=0}^{\infty} \phi_1^j \epsilon_{T+h-j} - \sum_{k=h}^{\infty} \phi_1^k \epsilon_{T+h-k} \\
e_T(h) &= \sum_{j=0}^{h-1} \phi_1^j \epsilon_{T+h-j}
\end{aligned}
$$

$$e_T(h) = \sum_{j=0}^{h-1} \phi_1^j \epsilon_{T+h-j}$$

The h-step ahead forecasts are unbiased, since

$$\mathbb{E}(e_T(h)) = \mathbb{E} \sum_{j=0}^{h-1} \phi_1^j \epsilon_{T+h-j} = \sum_{j=0}^{h-1} \phi_1^j \underbrace{\mathbb{E}(\epsilon_{T+h-j})}_{0} = 0$$

The forecast error variance is

$$\begin{aligned}
\mathrm{var}(e_T(h)) &= \mathrm{var} \sum_{j=0}^{h-1} \phi_1^j \epsilon_{T+h-j} \\
&= \sum_{j=0}^{h-1} \phi_1^{2j} \,\mathrm{var}(\epsilon_{T+h-j}) && [\,\{\epsilon_t\} \text{ is serially uncorrelated}] \\
&= \sigma^2 \sum_{j=0}^{h-1} \phi_1^{2j} && [\mathrm{var}(\epsilon_t) = \sigma^2, \forall t]
\end{aligned}$$

$$\begin{aligned} \mathrm{var}(e_T(h)) &= \sigma^2 \sum_{j=0}^{h-1} \phi_1^{2j} \\ &= \sigma^2 \frac{1-\phi_1^{2h}}{1-\phi_1^2} \qquad \text{[geometric series]} \end{aligned}$$

Note, since $\{Y_t\}$ is stationary, $|\phi_1| < 1$, and $\mathrm{var}(e_T(h))$ increases as h increases. In the limit:

$$\lim_{h\to\infty} \mathrm{var}(e_T(h)) = \frac{\sigma^2}{1-\phi_1^2} = \mathrm{var}(Y)\,.$$

Remark 10 (As forecast horizon $h \to \infty$)

- *forecasts tend to mean of process*
- *forecasts error variance tends to variance of process*

I.e. a finite data set cannot help forecast a long way into future

Example 11 (On board)

Consider problem of forecasting $MA(1)$ process, with non-zero mean:

$$Y_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1}\,.$$

Example 12 (One-step ahead forecast)

$$\begin{aligned}
\widehat{Y}_{T+1} &= \mathbb{E}(Y_{T+1}|Y_T,\dots,Y_1) \\
&= \mathbb{E}(\mu + \epsilon_{T+1} - \theta_1\epsilon_T|Y_T,\dots,Y_1) && \textit{[model equation]} \\
&= \mu + \overset{0}{\mathbb{E}(\epsilon_{T+1}|Y_T,\dots,Y_1)} - \theta_1\mathbb{E}(\epsilon_T|Y_T,\dots,Y_1) && [\epsilon_{T+1} \perp Y_{1:T}]^{a} \\
&= \mu - \theta_1\epsilon_T && \textit{[assuming we knew } \epsilon_T\textit{]}^{b}
\end{aligned}$$

I.e., one-step ahead prediction error is

$$\begin{aligned}
e_T(1) &= Y_{T+1} - \widehat{Y}_{T+1} \\
&= (\mu + \epsilon_{T+1} - \theta_1\epsilon_T) - (\mu - \theta_1\epsilon_T) && \textit{[model eqn. and above result]} \\
&= \epsilon_{T+1}\,.
\end{aligned}$$

[a]Define $Y_{1:T} := Y_1,\dots,Y_T$

[b]Explained later

$$e_T(1) = \epsilon_{T+1}\,.$$

Hence, for $MA(1)$ model, expected one-step ahead forecast error is

$$\mathbb{E}(e_T(1)) = \mathbb{E}(\epsilon_{T+1}) = 0\,,$$

and variance of one-step ahead forecast error is

$$\mathrm{var}(e_T(1)) = \mathrm{var}(\epsilon_{T+1}) = \sigma^2\,.$$

Example 13 ($h \geq 2$-step ahead forecast)

$$\begin{aligned}
\widehat{Y}_{T+h} &= \mathbb{E}(Y_{T+h}|Y_T,\ldots,Y_1) \\
&= \mathbb{E}(\mu + \epsilon_{T+h} - \theta_1\epsilon_{T+h-1}|Y_T,\ldots,Y_1) \qquad \textit{[model equation]} \\
&= \mu + \cancelto{0}{\mathbb{E}(\epsilon_{T+h}|Y_T,\ldots,Y_1)} - \theta_1\cancelto{0,\,h\geq 2}{\mathbb{E}(\epsilon_{T+h-1}|Y_T,\ldots,Y_1)} \\
&= \mu\,,
\end{aligned}$$

where the $\mathbb{E}(\cdot|Y_{1:T}) = 0$ *are due to* $\epsilon_{T+h} \perp Y_{1:T}$ *for* $h \geq 1$ *and* $\epsilon_{T+h-1} \perp Y_{1:T}$, *for* $h \geq 2$.

$$\widehat{Y}_{T+h} = \mu\,.$$

Hence, for $MA(1)$ model, the h-step ahead forecast error is

$$e_T(h) = Y_{T+h} - \widehat{Y}_{T+h} = \cancel{\mu} + \epsilon_{T+h} - \theta_1 \epsilon_{T+h-1} - \cancel{\mu}\,.$$

The mean h-step ahead forecast error is

$$\mathbb{E}(e_T(h)) = \mathbb{E}(\epsilon_{T+h} - \theta_1 \epsilon_{T+h-1}) = 0\,,$$

and variance of h-step ahead forecast error is

$$\text{var}(e_T(h)) = \text{var}(\epsilon_{T+h} - \theta_1 \epsilon_{T+h-1}) = \text{var}(Y_{T+h}) = \text{var}(Y)\,.$$

Remark 14 (Summary, $MA(1)$ forecasts)

	$h = 1$	$h \geq 2$
$\widehat{Y}_{T+h}$	$\mu - \theta_1 \epsilon_T$	μ
$e_T(h)$	ϵ_{T+1}	$\epsilon_{T+h} - \theta_1 \epsilon_{T+h-1}$
$\mathbb{E}(e_T(h)$	0	0
$\text{var}(e_T(h))$	σ^2	$\text{var}(Y)$

Proposition 15 (recall footnote in Example 12)

Consider **invertible** $MA(1)$ *model. For* **large** T*:*

$$\mathbb{E}(\epsilon_T | Y_T, \ldots, Y_1) = \epsilon_T \,.$$

'Proof' Consider zero-mean $MA(1)$ process, $Y_t = (1 - \theta_1 B)\epsilon_t$. Since $\{Y_t\}$ is invertible, we have, at time T:

$$\epsilon_T = (1 - \theta_1 B)^{-1} Y_T = \sum_{j=0}^{\infty} \theta_1^j Y_{T-j} \,.$$

I.e., given $Y_T, Y_{T-1}, \ldots$, for large T, we can find $\epsilon_T, \epsilon_{T-1}, \ldots$. Hence

$$\mathbb{E}(\epsilon_T | Y_T, Y_{T-1}, \ldots) = \epsilon_T \,, \qquad T \text{ large.}$$

I.e. expected observed white noise is the observed white noise (!); expected unobserved (future) white noise is zero.

In practice, assuming correct model and good estimates (large T), the residuals $\varepsilon_T := Y_{T+1} - \mathbb{E}(Y_{T+1} | Y_{1:T})$ can be used to approximate the required white noise terms:

$$\varepsilon_T \approx \epsilon_T \,.$$

Consider stationary and invertible, zero-mean $ARMA(p,q)$ process:

$$Y_t = \sum_{j=1}^{p} \phi_j Y_{t-j} + \epsilon_t + \sum_{k=1}^{q} \theta_k \epsilon_{t-k} \, .$$

The h-step ahead forecast, with origin T is

$$\begin{aligned}
\widehat{Y}_{T+h} &= \mathbb{E}(Y_{T+h} | Y_T, \ldots, Y_1) \qquad\qquad [\mathit{define}\, Y_{1:T} := Y_1, \ldots, Y_T] \\
&= \mathbb{E}\left(\sum_{j=1}^{p} \phi_j Y_{T+h-j} + \epsilon_{T+h} - \sum_{k=1}^{q} \theta_k \epsilon_{T+h-k} \middle| Y_{1:T} \right) \\
&= \sum_{j=1}^{p} \phi_j \mathbb{E}(Y_{T+h-j} | Y_{1:T}) + \cancelto{0}{\mathbb{E}(\epsilon_{T+h} | Y_{1:T})} - \sum_{k=1}^{q} \theta_k \mathbb{E}(\epsilon_{T+h-k} | Y_{1:T}) \\
&= \sum_{j=1}^{p} \phi_j \widehat{Y}_{T+h-j} - \sum_{k=1}^{q} \theta_k \mathbb{E}(\epsilon_{T+h-k} | Y_{1:T})
\end{aligned}$$

$$\widehat{Y}_{T+h} = \sum_{j=1}^{p} \phi_j \widehat{Y}_{T+h-j} - \sum_{k=1}^{q} \theta_k \mathbb{E}(\epsilon_{T+h-k} | Y_{1:T})$$

Now, note that,

Corollary 16

From Proposition 15 (unobserved [future] white noise has zero-mean):

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

I.e.

$$\widehat{Y}_{T+h} = \sum_{j=1}^{p} \phi_j \widehat{Y}_{T+h-j} - \sum_{k=h}^{q} \theta_q \epsilon_{T+h-k} \,.$$

Remark 17

To investigate forecast error, it is 'convenient' to express the model as an MA process.

Theorem 18 (Recall lecture 4)

Let $\{Y_t\}$ *be a stationary* $ARMA(p,q)$ *process, with* $\mathbb{E}(Y_t)=\mu$, $AR(p)$ *characteristic polynomial*

$$\phi(B) := 1 - \sum_{j=1}^{p} \phi_j B^j \,,$$

and $MA(q)$ *characteristic polynomial*

$$\theta(B) := 1 - \sum_{j=1}^{p} \theta_j B^j \,.$$

I.e. $\phi(B)(Y_t-\mu)=\theta(B)\epsilon_t$. *Then*

$$\exists\, \psi(B) = \frac{\theta(B)}{\phi(B)} =: \sum_{j=0}^{\infty} \psi_j B^j \,,$$

such that

$$Y_t - \mu = \psi(B)\epsilon_t \,.$$

Corollary 19

Any stationary $ARMA(p,q)$ process can be written as

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

Proof (Trivial?) By definition, $\phi(B)(Y_t - \mu) = \theta(B)\epsilon_t$. Hence,

$$Y_t - \mu = \frac{\theta(B)}{\phi(B)}\epsilon_t = \psi(B)\epsilon_t = \sum_{j=0}^{\infty} \psi_j B^j \epsilon_t \, . \qquad \blacksquare$$

Corollary 20

Let Y_t be a stationary and invertible $ARMA(p,q)$ process. Then the h-step ahead forecast error, with origin T, satisfies:

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

Proof Since $\{Y_t\}$ stationary, we have, from Corollary 19,

$$Y_{T+h} = \mu + \sum_{j=0}^{\infty} \psi_j \, \epsilon_{T+h-j} \, .$$

The $h-$step ahead forecast with origin T is

$$\begin{aligned}
\widehat{Y}_{T+h} &= \mathbb{E}(Y_{T+h} | Y_{1:T}) \\
&= \mathbb{E}\left(\mu + \sum_{j=0}^{\infty} \psi_j \, \epsilon_{T+h-j} \middle| Y_{1:T}\right) \\
&= \mu + \sum_{j=0}^{\infty} \psi_j \, \mathbb{E}(\epsilon_{T+h-j} | Y_{1:T}) \\
&= \mu + \sum_{j=h}^{\infty} \psi_j \, \epsilon_{T+h-j} \, ,
\end{aligned}$$

since we have (for invertible $\{Y_t\}$), from Corollary 16 that:

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

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

Therefore, for any stationary, invertible $ARMA(p, q)$ process:

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

Also,

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