Exercise Sheet 6

$\boxed{1}$ a) No, the MSE is scale-dependent. Take the following VAR(1) as example ($\mu_z = 0$ w.l.o.g.):

$$z_t = \phi_1 z_{t-1} + a_t$$

Define $k_t := b z_t$ when b is a scalar.

$$\Rightarrow \quad k_t = b z_t = \phi_1 b z_{t-1} + e_t$$

$$\Leftrightarrow \quad \underline{e_t} = k_t - \phi_1 k_t = b \cdot (z_t - \phi_1 z_{t-1})$$

$$= \underline{b \cdot a_t}$$

$$MSE(\hat{k}_{t,t+1}) = \mathbb{E}(e_{t+1}^2) = \mathbb{E}((b a_{t+1})^2)$$

$$= b^2 \cdot \mathbb{E}(a_{t+1}^2)$$

$$= b^2 \cdot MSE(\hat{z}_{t,t+1})$$

where $\hat{k}_{t,t+1}, \hat{z}_{t,t+1}$ are the VAR(1) predictions.

b) $$\mathcal{X}(\ell) = \underbrace{\log(|\hat{\Sigma}_a(\ell)|)}_{\substack{\text{Fit} \\ \sim \log(|MSE|)}} + \underbrace{\frac{\ell}{T} \cdot c_T}_{\text{Complexity}}$$

c) Again, the VAR(1) example.

$$z_t = \phi_0 + \phi_1 z_{t-1} + a_t \quad \Big| \; z_t \text{ is } k \times 1$$

Linear transformation: $k_t = B \cdot z_t + c$

$\Rightarrow$ c is covered by $\tilde{\phi}_0 = \phi_0 + c$, no problem.

w.l.o.g. we omit that part

$$\Rightarrow \underbrace{B z_t}_{=: k_t} = \phi_1 \underbrace{B z_{t-1}}_{=: k_{t-1}} + \underbrace{B a_t}_{=: e_t}$$

$$\Rightarrow \left| MSE(\hat{k}_{t,t+1}) \right| = \left| \mathbb{E}(e_t e_t') \right|$$

$$= \left| \mathbb{E}\left(\underbrace{\boxed{B}}_{K\times K} \underbrace{\boxed{a_t a_t'}}_{K\times K} \underbrace{\boxed{B'}}_{K\times K} \right) \right|$$

$$= \left| B \cdot \underbrace{\mathbb{E}(a_t a_t')}_{= MSE(\hat{\Sigma}_{t,t+1})} \cdot B' \right|$$

$$= |B| \cdot |MSE(\hat{\Sigma}_{t,t+1})| \, |B'|$$

$$= (|B|)^2 \cdot |MSE(\hat{\Sigma}_{t,t+1})|$$

$\Rightarrow$ The linear transformation affects the value of the ICs! But as long as $|B| \neq 0$ (non-singular), the $MSE(\hat{K}_{t,t+1})$ is minimal when

$MSE(\hat{z}_{t|t+1})$ has its minimum.

Example for singular B:

$$B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 8 \\ 0 & 0 & 8 \\ 0 & 0 & 1 \end{pmatrix}$$

d) In (3-3), we write the VAR(p) model as: $Z = X\beta + A$

$$\Leftrightarrow A = Z - X\beta$$

and $\hat{\beta} = (X'X)^{-1}X'Z = \beta + (X'X)^{-1}X'A$

$$\Rightarrow Var(\hat{\beta} - \beta) = \mathbb{E}\left((X'X)^{-1}X'A \left[(X'X)^{-1}X'A\right]'\right)$$

$$= \mathbb{E}\left((X'X)^{-1}X'AA'X \quad (X'X)^{-1}\right)$$

If we scale Z by the scalar b, we also scale $X = (LZ, L^2 Z, \ldots)$ and A by b. $\Rightarrow \tilde{X} = bX$, $\tilde{Z} = bZ$, $\tilde{A} = bA$, $\tilde{\beta} = \frac{b^2}{b^2}\beta$

$$\Rightarrow Var(\hat{\tilde{\beta}} - \beta) = \mathbb{E}\left((\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{A}\tilde{A}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\right)$$

$$= \mathbb{E}\left(\frac{1}{b^2}(X'X)^{-1}\, b^2\, X'AA'X\, b^2\, \frac{1}{b^2}(X'X)^{-1}\right)$$

$$= Var(\hat{\beta} - \beta)$$

$\Rightarrow$ Standard errors are scale-invariant! (similar to R^2).

$\boxed{2}$ a) HQ = 4

b) 78

c) i) 71 ii) No. But close.

iii) Separate tests give 78, but the joint (multiple) test gives 71. Most coefficients initially explained tiny bits of the variation. And if these coefficients are correlated with each other, restricting some coefficients changes the remaining coefficients, forcing an earlier rejection.

→ Problems in backward selection!

d) The refined model wins, all 3 ICs support it. Also note how close the ICs' values are in comparison to the fully specified model.

e) No, the fully specified model performs better. It is more complex and can therefore model complex dynamics better (in-sample). But it is questionable whether those dynamics are deterministic or just noise ($\leftrightarrow$ overfitting).

f) $\# \text{coefs} = K + \underline{K^2 \cdot p}$

Note that we do not adjust for intercept!
(since demeaning would be needed anyways!)

i) Ljung-Box test rejects H_0 everywhere:
$\Rightarrow$ dynamic pattern in the residuals
$\Rightarrow$ VAR(4) did not absorb everything of it, but we did not expect this after last week.

ii) The full model uses $4 \cdot 25 = 100$ coefficients (without intercept!), the refined model just 32
(equivalent to a fully specified model with $p = 1.28$).

Since $m > p$ is required for the Ljung-Box test, there is no result for $m \leq p$. (see slide 4-14 for the test)

g) The VAR(4) is also better than a VAR(1) MSE-wise.
(not surprising at in-sample, see e))

h) i) At $h=1$, the VAR(4) is better. ($h=1$ corresponds to the in-sample MSE, so this is hardly surprising).

But as $h \geq 2$, the Sparser VAR(1) model fairs much better, because it is less prone to overfitting which hurts the out-of-sample forecasts.

$\Rightarrow$ Overfitting: Adding a lot of coefficients makes the model explain even the "unexplainable" (i.e. stochastic) parts, resulting in better in-sample but worse out-of-sample predictions. Such models are "overspecified".

ii) MSE: in-sample prediction errors

MSFE: out-of-sample prediction errors

iii) They converge to the means $\mathbb{E}(z_t)$, because this process is stationary and the influence of $a_t, a_{t-1}, \ldots, a_0$ vanishes if $h \to \infty$.

$\boxed{3}$ a) $\rightarrow$ R

b) All.

c) Virtually the same. Though the VAR(1) explained tiny bits of the variation.

d) VAR(0) with intercept:

$$z_t = \phi_0 + a_t$$

$$\Rightarrow \hat{\phi}_0 = \arg\min_{\phi_0} \sum_{t=1}^{T} (z_t - \phi_0)'(z_t - \phi_0)$$

$$\Rightarrow \hat{\phi}_0 = \frac{1}{T} \sum_{t=1}^{T} z_t \rightarrow \mathbb{E}(z_t)$$

$$\Rightarrow \hat{z}_{t,t+h} = \overline{z_t} = \hat{\phi}_0$$

The VAR(1) does slightly better at $h = \{1,2,3\}$ but then the forecast errors align because the VAR(1) forecast has reverted to the mean.

e) The VAR(1) might be overspecified, but this did not lead to considerable overfitting. Primarily this is due to the small number of coefficients and the large sample size ($\frac{\#\text{coefs}}{\#\text{data points}}$ remains small)

f) Raise $\frac{\#\text{ coefs}}{\#\text{ data points}}$

$T \downarrow$, $p \uparrow$

$\rightarrow R$

$\boxed{4}$ To show:

The elements of the optimal <u>multivariate</u> forecasts constitute optimal <u>univariate</u> forecasts.

$\hat{Z}^{(i)}_{T,T+h}$: h-steps ahead forecast of variable 'i' at time T

$Z^{(i)}_{T}(h)$: Optimal h-steps ahead forecast of variable 'i' at time T

<u>Multivariate</u>: Equation 5.1

$$|MSE(\hat{Z}_{T,T+h})| \geq |MSE(Z_T(h))|$$

$$\Leftrightarrow |\underbrace{MSE(\hat{Z}_{T,T+h}) - MSE(Z_T(h))}_{\to \text{ a p.s.d. matrix!}}| \geq 0$$

From slide (5-5) we know:

$$= \mathbb{E}\left((Z_T(h) - \hat{Z}_{T,T+h})(Z_T(h) - \hat{Z}_{T,T+h})' \right)$$

$=: A$

For any p.s.d. matrix A and vector ω it holds that

$$\omega' A \omega \geq 0 \quad \left(\begin{matrix}\text{quadratic} \\ \text{form}\end{matrix}\right)$$

Define $\omega = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow$ only 1 at index 'i' !

$\Rightarrow$ $\omega' A$ is 0 everywhere except at row 'i' and $A\omega$ sets every column except 'i' to zero.

$\rightarrow$ $\omega' A \omega$ is only $\neq 0$ at the ith element on the main diagonal!

That means:

$$w' A w = \mathbb{E}\left(\left(Z_T^{(i)}(h) - \hat{Z}_{T,T+h}^{(i)}\right)^2\right)$$

$$\geq 0 \quad (\text{p.s.d.})$$

which corresponds to

$$MSE\left(\hat{Z}_{T,T+h}^{(i)}\right) - MSE\left(Z_T^{(i)}(h)\right)$$

$$= \mathbb{E}\left(\left(\hat{Z}_{T,T+h}^{(i)} - Z_{T+h}^{(i)}\right)^2\right) - \mathbb{E}\left(\left(Z_T^{(i)}(h) - Z_{T+h}^{(i)}\right)^2\right)$$

$$\geq 0$$

Reminder: $MSE(\hat{z}_{T,T+h})$

$$
= \begin{pmatrix}
(\hat{z}^{(1)}_{T,T+h} - z^{(1)}_{T+h})^2 & (\hat{z}^{(1)}_{T,T+h} - z^{(1)}_{T+h})(\hat{z}^{(2)}_{T,T+h} - z^{(2)}_{T+h}) & \cdots \\
(\hat{z}^{(2)}_{T,T+h} - z^{(2)}_{T+h})(\hat{z}^{(1)}_{T,T+h} - z^{(1)}_{T+h}) & (\hat{z}^{(2)}_{T,T+h} - z^{(2)}_{T+h})^2 & \\
\vdots & & \ddots \\
(\hat{z}^{(k)}_{T,T+h} - z_{T+h})(\hat{z}^{(1)}_{T,T+h} - z^{(1)}_{T+h}) & & (\hat{z}^{(k)}_{T,T+h} - z_{T+h})^2
\end{pmatrix}
$$

Keep mind that $|(A-B)| \geq 0$

does <u>not</u> imply that $a_{ii} - b_{ii} \geq 0 \quad \forall i$

in general!