RESIT TERM 2021

24-HOUR ONLINE EXAMINATION

ECON0058: TIME SERIES ANALYSIS

All work must be submitted anonymously. Please ensure that you add your **candidate number** and the module code to the template answer sheet provided. Note that the candidate number is a combination of four letters plus a number, e.g. ABCD9. You can find your candidate number in your PORTICO account, under "My Studies" then the "Examinations" container. Please, note that the candidate number is NOT the same as your student number (8 digits), which is printed on your UCL ID card. Submitting with your student number will delay marking and when your results might be available.

Word count: 2000
The above word count is provided as guidance only on the expected total length of your submitted answer sheets. You will not be penalised if you exceed the word limit.

*Answer **ALL FOUR** questions*

Each of the FOUR questions carries the same weight (25 points)

Allow enough time to submit your work. Waiting until the deadline for submission risks facing technical problems when submitting your work, due to limited network or systems capacity.

*By submitting this assessment, **you pledge your honour** that you have not violated UCL's Assessment Regulations which are detailed in* https://www.ucl.ac.uk/academic-manual/chapters/chapter-6-student-casework-framework/section-9-student-academic-misconduct-procedure, *which include (but are not limited to) plagiarism, self-plagiarism, unauthorised collaboration between students, sharing my assessment with another student or third party, access another student's assessment, falsification, contract cheating, and falsification of extenuating circumstances.*

Question 1

Consider the AR(2) process

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t,\ \varepsilon_t \sim iid(0, \sigma^2)$$

(a) [5 points] Derive the optimal 2-step-ahead forecast for a quadratic loss, and show how it is related to the optimal 1-step-ahead forecast.

(b) [5 points] Derive the optimal 2-step-ahead forecast error. What time series model describes its behavior?

(c) [5 points] Derive the variance of the 1-step ahead and the 2-step ahead forecast.

(d) [5 points] Derive the h-step ahead optimal forecast for a quadratic loss. What is the long run forecast, i.e. the forecast for an horizon h going to infinity?

(e) [5 points] Consider the task of producing a time series of realized forecast errors. Describe in detail how this be achieved in practice. Describe and discuss the advantages and disadvantages of a recursive versus a rolling estimation scheme for such exercise.

Question 2

Define $f_{t+\tau|t}$ as the forecast of the variable $Y_{t+\tau}$ a the $\tau-$step-ahead horizon, based on the information set Ω_t. An economist wishes to produce optimal forecasts for a quadratic loss function and evaluate their optimality.

(a) [5 points] Suppose $Y_t = 0.5Y_{t-1} + 0.1Y_{t-2} + \varepsilon_t + 0.5\varepsilon_{t-1} + 0.25\varepsilon_{t-2}$. Derive the optimal one-step ahead forecast and corresponding forecast error for a quadratic loss.

(b) [5 points] Suppose $Y_t = 0.5Y_{t-1} + 0.1Y_{t-2} + \varepsilon_t + 0.5\varepsilon_{t-1} + 0.25\varepsilon_{t-2}$. Derive the optimal $\tau-$step-ahead forecast and corresponding forecast error for a quadratic loss, for $\tau \geq 2$.

(c) [5 points] Consider the regression

$$e_{t+\tau|t} = \alpha + \eta_t, \tag{1}$$

  (a) where $e_{t+\tau|t} = Y_{t+\tau} - f_{t+\tau|t}$ are the $\tau-$step-ahead forecast errors and η_t is an error term.

  (i) What is the implication of $f_{t+\tau|t}$ being the best forecast (for a quadratic loss function) on the coefficient α in the regression (1)?

  (ii) How would you test the for this implication?

  (iii) Which estimator for the variance of $\hat{\alpha}$ would you use in a t-test on this coefficient? Why?

(d) [5 points] Now consider the following *alternative* loss function: $L(e_{t+\tau|t}) = \exp(e_{t+\tau|t}) - e_{t+\tau|t} - 1$. Explain how you would modify equation (1) to test forecast optimality for this alternative loss function. [10 points]

(e) [5 points] Consider the task of producing a time series of realized forecast errors $e_{t+\tau|t}$ to use in equation (1). Describe in detail how this can be achieved in practice. Describe and discuss the advantages and disadvantages of a recursive versus a rolling estimation scheme for such exercise.

Question 3

Consider the following GARCH(1,2) model for the stock returns r_t:

$$\begin{aligned} r_t &= \mu + \eta_t; \\ \eta_t &= h_t \varepsilon_t; \ \varepsilon_t \sim iidN(0,1), \\ h_t^2 &= \alpha_0 + \alpha_1 \eta_{t-1}^2 + \beta_1 h_{t-1}^2 + \beta_2 h_{t-2}^2; \end{aligned} \tag{2}$$

where r_t is the return on a stock. Assume that all necessary restrictions hold on the coefficients so that the process is stationary and h_t^2 is always non-negative. Furthermore, assume that the fourth moment $E[\eta_t^4] = m_4$ exists and does not depend on time. Answer all the following questions.

(a) [5 points] Compute the conditional mean and the conditional variance of r_t, conditioning on the information available in period $t-1$.

(b) [5 points] Show that η_t^2 follows an ARMA(1,2) model, while η_t is not autocorrelated. Which stylized fact in the data is consistent with this finding? [10 points]

(c) [5 points] Show that η_t features excess kurtosis, i.e.

$$\kappa = \frac{E[\eta_t^4]}{E[\eta_t^2]^2} > 3.$$

Which stylized fact in the data is consistent with this finding? [10 points]

(d) [5 points] Set $\beta_2 = 0$. Derive the fourth moment $E[\eta_t^4]$ as a function of α_1 and β_1. Show that a necessary condition for the fourth moment to exist is $1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2 > 0$ [10 points]

(e) [5 points] Set $\beta_2 = 0$ and $1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2 > 0$. Derive the kurtosis κ as a function of α_1 and β_1. Show that the kurtosis κ exceeds the kurtosis of ε_t by the quantity $6\alpha_1^2/\{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2\}$. [10 points]

Question 4

Consider the following data generating process:

$$
\begin{aligned}
x_t &= u_{1t} && (3)\\
y_t &= \alpha x_t + u_{2t} && (4)
\end{aligned}
$$

Where $u_{1t} = \theta u_{1t-1} + \epsilon_{1t}, u_{2t} = \rho u_{2t-1} + \epsilon_{2t},\ E[\epsilon_{1t}] = E[\epsilon_{2t}] = 0,\ Var[\epsilon_{it}] = \sigma^2, \forall i = 1, 2$ and $Cov[\epsilon_{1t}, \epsilon_{2t}] = \gamma$. The shocks $(\epsilon_{1t}, \epsilon_{2t})$ are the fundamental shocks of the Wold representation.

(a) [5 points] Derive the degree of integratedness of the two series, x_t and y_t, considering the cases $|\theta| < 1$ and $\theta = 1$, $|\rho| < 1$ and $\rho = 1$, and for $\alpha = 0$ and $\alpha \neq 0$. Under what coefficient restrictions are x_t and y_t cointegrated? What is the cointegrating vector in such cases?

(b) [5 points] Derive the Vector Auto Regression representation:

$$
\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \theta & 0 \\ \alpha(\theta - \rho) & \rho \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \quad (5)
$$

HINT: start by solving (3) and (4) for u_{1t} and u_{2t}.

(c) [5 points] Compute the roots of the characteristic polynomial of (5). By making the appropriate assumptions on these roots, derive the Vector Moving Average representation of the process.

(d) [5 points] By making the appropriate assumptions on the coefficients, derive the Vector Error Correction (VEC) representation from (5).

(e) [5 points] Discuss the advantages/disadvantages of imposing the VEC representation on the data, as opposed to using the VAR in levels shown in (5)

Solution to question 1

Consider the AR(2) process $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t,\ \varepsilon_t \sim iid(0, \sigma^2)$

1. The optimal 2-step-ahead forecast for a quadratic loss is $f^*_{t+2|t} = E(Y_{t+2}|\Omega_t)$. We have $Y_{t+2} = \phi_1 Y_{t+1} + \phi_2 Y_t + \varepsilon_{t+2}$, so the optimal forecast is

$$f^*_{t+2|t} = E(Y_{t+2}|\Omega_t) = \phi_1 E(Y_{t+1}|\Omega_t) + \phi_2 Y_t, \tag{6}$$

   and the forecast error is ε_{t+2} (notice how this is different from the case of an MA(2), where $Y_{t+2} = \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1} + \theta_2 \varepsilon_t$, and $E(Y_{t+2}|\Omega_t) = \theta_2 \varepsilon_t$ since ε_{t+1} is not known at time t, i.e., $E(\varepsilon_{t+1}|\Omega_t) = 0$). This formula shows that the 2-step-ahead forecast is a linear combination of the realization of the variable at time t, Y_t, and of the 1-step-ahead forecast, $f^*_{t+1|t} = E(Y_{t+1}|\Omega_t)$. We can then derive the expression for $E(Y_{t+1}|\Omega_t)$ by writing $Y_{t+1} = \phi_1 Y_t + \phi_2 Y_{t-1} + \varepsilon_{t+1}$, which gives $E(Y_{t+1}|\Omega_t) = \phi_1 Y_t + \phi_2 Y_{t-1}$ and substitute it into (6) to obtain $f^*_{t+2|t} = E(Y_{t+2}|\Omega_t) = \left(\phi_1^2 + \phi_2\right) Y_t + \phi_1 \phi_2 Y_{t-1}$.

   (a) An alternative way to derive the optimal forecast, which is useful to answer the next question, is to use recursive substitution, i.e., to write $Y_{t+2} = \phi_1 Y_{t+1} + \phi_2 Y_t + \varepsilon_{t+2} = \phi_1(\phi_1 Y_t + \phi_2 Y_{t-1} + \varepsilon_{t+1}) + \phi_2 Y_t + \varepsilon_{t+2}$. From this, it is easy to see that $f^*_{t+2|t} = E(Y_{t+2}|\Omega_t) = E(\left(\phi_1^2 + \phi_2\right) Y_t + \phi_1 \phi_2 Y_{t-1} + \varepsilon_{t+2} + \phi_1 \varepsilon_{t+1}|\Omega_t)$, which, as before, gives $f^*_{t+2|t} = \left(\phi_1^2 + \phi_2\right) Y_t + \phi_1 \phi_2 Y_{t-1}$

2. We have $e_{t+2|t} = Y_{t+2} - f^*_{t+2|t}$. Using recursive substitution, as in the last part of answer a., we can write $Y_{t+2} = \left(\phi_1^2 + \phi_2\right) Y_t + \phi_1 \phi_2 Y_{t-1} + \varepsilon_{t+2} + \phi_1 \varepsilon_{t+1}$. Therefore, we have $e_{t+2|t} = \varepsilon_{t+2} + \phi_1 \varepsilon_{t+1}$, which is an MA(1).

3. The 1-step-ahead forecast variance is

$$V(Y_{t+1}|\Omega_t) = V(\phi_1 Y_t + \phi_2 Y_{t-1} + \varepsilon_{t+1}|\Omega_t) = V(\varepsilon_{t+1}|\Omega_t) = \sigma^2$$

   The 2-step-ahead forecast variance is

$$\begin{aligned} V(Y_{t+2}|\Omega_t) &= V(\phi_1 Y_{t+1} + \phi_2 Y_t + \varepsilon_{t+2}) \\ &= \phi_1^2 V(Y_{t+1}|\Omega_t) + V(\varepsilon_{t+2}|\Omega_t) \\ &= \phi_1^2 V(\varepsilon_{t+1}|\Omega_t) + V(\varepsilon_{t+2}|\Omega_t) \\ &= (\phi_1^2 + 1)\sigma^2 \end{aligned}$$

4. The easiest way to solve this is to rewrite the process as follows:

$$\begin{aligned} \begin{bmatrix} Y_t \\ Y_{t-1} \end{bmatrix} &= \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix} \\ z_t &= \Phi z_{t-1} + v_t \end{aligned}$$

$$\hat{z}_{t+h} = \Phi^h z_t$$

   Since for h going to infinity Φ^h goes to zero, the long run forecast coincides with the unconditional mean of the process, and it is 0

5. The discussion should be based on the lecture notes and slides. Students are expected to describe how an out of sample forecasting exercises performed. Moreover, they need to stress the fact that a recursive scheme expands the estimation window while the rolling scheme keeps the sample size constant. The former is more efficient, but the latter is robust to structural breaks and changes in regime. There are other ways to prove this result, involving more convoluted expressions

Solution to question 2

1. (a) Consider the general model

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}.$$

The optimal forecast for a quadratic loss is the conditional mean, and thus

$$\begin{aligned} f_{t+1|t} &= E(Y_{t+1}|\Omega_t) = E(\phi_1 Y_t + \phi_2 Y_{t-1} + \varepsilon_{t+1} + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1}|\Omega_t) = \phi_1 Y_t + \phi_2 Y_{t-1} + \theta_1 \varepsilon \\ e_{t+1|t} &= Y_{t+1} - f_{t+1|t} = \varepsilon_{t+1} \end{aligned}$$

Using the values of the specific model we have

$$\begin{aligned} f_{t+1|t} &= 0.5Y_t + 0.1Y_{t-1} + 0.5\varepsilon_t + 0.25\varepsilon_{t-1} \\ e_{t+1|t} &= Y_{t+1} - f_{t+1|t} = \varepsilon_{t+1} \end{aligned}$$

(b) The optimal forecast for a quadratic loss is the conditional mean. This is an ARMA (2,2) and therefore the forecast will be similar to that on an AR model after adjusting for the impact of the MA component. The MA component will only play a role for the 2-step ahead case, since for every step beyond that there is no information in the MA component. Hence we have, for $\tau = 2$:

$$\begin{aligned} f_{t+2|t} &= E(Y_{t+2}|\Omega_t) = E(\phi_1 Y_{t+1} + \phi_2 Y_t + \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1} + \theta_2 \varepsilon_t|\Omega_t) \\ &= \phi_1 f_{t+1|t} + \phi_2 Y_t + \theta_2 \varepsilon_t \\ e_{t+2|t} &= Y_{t+2} - f_{t+2|t} \\ &= \phi_1 Y_{t+1} + \phi_2 Y_t + \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1} + \theta_2 \varepsilon_t - (\phi_1 f_{t+1|t} + \phi_2 Y_t + \theta_2 \varepsilon_t) \\ &= \phi_1 Y_{t+1} - \phi_1 f_{t+1|t} + \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1} \\ &= \phi_1 \varepsilon_{t+1} + \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1} = (\phi_1 + \theta_1)\varepsilon_{t+1} + \varepsilon_{t+2} \end{aligned}$$

Using the values of the specific model we have

$$\begin{aligned} f_{t+2|t} &= 0.5 f_{t+1|t} + 0.1Y_t + 0.25\varepsilon_t \\ e_{t+2|t} &= \varepsilon_{t+1} + \varepsilon_{t+2} \end{aligned}$$

For $\tau > 2$:

$$\begin{aligned} f_{t+\tau|t} &= E(Y_{t+\tau}|\Omega_t) = E(\phi_1 Y_{t+\tau-1} + \phi_2 Y_{t+\tau-2} + \varepsilon_{t+\tau} + \theta_1 \varepsilon_{t+\tau-1} + \theta_2 \varepsilon_{t+\tau-2}|\Omega_t) \\ &= \phi_1 f_{t+\tau-1} + \phi_2 f_{t+\tau-2} \\ e_{t+\tau|t} &= Y_{t+\tau} - f_{t+\tau|t} \\ &= \phi_1 Y_{t+\tau-1} + \phi_2 Y_{t+\tau-2} + \varepsilon_{t+\tau} + \theta_1 \varepsilon_{t+\tau-1} + \theta_2 \varepsilon_{t+\tau-2} - (\phi_1 f_{t+\tau-1} + \phi_2 f_{t+\tau-2}) \\ &= \phi_1 (Y_{t+\tau-1} - f_{t+\tau-1}) + \phi_2 (Y_{t+\tau-2} - f_{t+\tau-2}) + \varepsilon_{t+\tau} + \theta_1 \varepsilon_{t+\tau-1} + \theta_2 \varepsilon_{t+\tau-2} \\ &= \varepsilon_{t+\tau} + (\theta_1 + \phi_1)\varepsilon_{t+\tau-1} + (\theta_2 + \phi_2)\varepsilon_{t+\tau-2} \end{aligned}$$

Using the values of the specific model we have

$$\begin{aligned} f_{t+\tau|t} &= 0.5f_{t+\tau-1} + 0.1f_{t+\tau-2} \\ e_{t+\tau|t} &= \varepsilon_{t+\tau} + \varepsilon_{t+\tau-1} + 0.35\varepsilon_{t+\tau-2} \end{aligned}$$

(c) Optimality for the quadratic loss $L(e_{t+\tau|t}) = e^2_{t+\tau|t}$ means that $E\left(\partial L/\partial e_{t+\tau|t}|\Omega_t\right) = 2E\left(e_{t+\tau|t}|\Omega_t\right) = 0$, which in turn implies that $\alpha = E\left(e_{t+\tau|t}\right) = 0$. To test this null, we would need a t-test with a HAC correction, since $\tau-$step ahead forecast errors are serially correlated for $\tau > 1$. Students are expected to provide plenty of details and discussion on this.

(d) Optimality for a linex loss means $E\left(\partial L/\partial e_{t+\tau|t}|\Omega_t\right) = E\left(\exp(e_{t+\tau|t}) - 1|\Omega_t\right) = 0$. You can thus create a new variable $Z_{t+\tau} = \exp(e_{t+\tau|t}) - 1$ and run the regression $Z_{t+\tau} = \alpha + X_t'\beta + \varepsilon_{t+\tau}$, where X_t is a set of variables from the information set at time t, Ω_t. Forecast optimality is tested by a joint test of the hypothesis that α and β equal 0. Also in this case a HAC correction is necessary.

2. Students are expected to describe how an out of sample forecasting exercise is performed. Moreover, they need to stress the fact that a recursive scheme expands the estimation window while the rolling scheme keeps the sample size constant. The former is more efficient, but the latter is robust to structural breaks and changes in regime.

Solution to question 3

1. The conditional expectations are:

$$
\begin{aligned}
E[r_t|\Omega_{t-1}] &= \mu + E[\eta_t|\Omega_{t-1}] = \mu + E[h_t\varepsilon_t|\Omega_{t-1}] = \mu + h_t E[\varepsilon_t|\Omega_{t-1}] = \mu \\
V[r_t|\Omega_{t-1}] &= V[\eta_t|\Omega_{t-1}] = V[\eta_t|\Omega_{t-1}] = E[\eta_t^2|\Omega_{t-1}] - E[\eta_t|\Omega_{t-1}]^2 = E[h_t^2\varepsilon_t^2|\Omega_{t-1}] = h_t^2
\end{aligned}
$$

2. Show that η_t^2 follows an ARMA(1,2) model, while η_t is not autocorrelated. Which stylized fact in the data is consistent with this finding?

   First, students should show that the autocovariance function of η_t is:

$$
\begin{aligned}
E[\eta_t^2] &= \sigma^2 \\
E[\eta_t\eta_{t-1}] &= 0 \\
&\dots \\
E[\eta_t\eta_{t-s}] &= 0
\end{aligned}
$$

   Then, they should show that:

$$
\begin{aligned}
E[\eta_t^2|\Omega_{t-1}] &= h_t^2 \\
\eta_t - E[\eta_t^2|\Omega_{t-1}] &= \eta_t^2 - h_t^2 \ = v_t \\
h_t^2 &= \alpha_0 + \alpha_1\eta_{t-1}^2 + \beta_1 h_{t-1}^2 + \beta_2 h_{t-2}^2 \\
\eta_t^2 - v_t &= \alpha_0 + \alpha_1\eta_{t-1}^2 + \beta_1(\eta_{t-1}^2 - v_{t-1}) + \beta_2(\eta_{t-2}^2 - v_{t-2}) \\
\eta_t^2 &= \alpha_0 + (\alpha_1+\beta_1)\eta_{t-1}^2 + \beta_2\eta_{t-2}^2 + v_t - \beta_1 v_{t-1} - \beta_2 v_{t-2}
\end{aligned}
$$

   and:

$$
\begin{aligned}
E[\eta_t^4] &= m_4 = E[\alpha_0 + (\alpha_1+\beta_1)\eta_{t-1}^2 + \beta_2\eta_{t-2}^2 + v_t - \beta_1 v_{t-1} - \beta_2 v_{t-2}] \\
E[\eta_t^2\eta_{t-1}^2] &= 0 \\
&\dots \\
E[\eta_t\eta_{t-s}] &= 0
\end{aligned}
$$

3. We have that

$$
\kappa(\eta_t) = \frac{E[\eta_t^4]}{E[\eta_t^2]^2} = \frac{E[h_t^4\varepsilon_t^4]}{E[h_t^2\varepsilon_t^2]^2} = \frac{E[h_t^4]E[\varepsilon_t^4]}{E[h_t^2]^2E[\varepsilon_t^2]^2} = \frac{E[\sigma_t^4]}{E[\sigma_t^2]^2}3 \geq 3
$$

   where the last inequality follows from Jensen's inequality $E[f(x)] \geq f(E[x])$.

4. The model simplifies to a GARCH(1,1) model:

$$\begin{aligned} r_t &= \mu + \eta_t; \\ \eta_t &= h_t \varepsilon_t; \quad \varepsilon_t \sim iidN(0,1) \\ h_t^2 &= \alpha_0 + \alpha_1 \eta_{t-1}^2 + \beta_1 h_{t-1}^2; \\ \alpha_0 &> 0; \alpha_1 \geq 0; \alpha_1 + \beta_1 < 1 \end{aligned}$$

Since ε_t is normal we have that

$$\eta_t | \Omega_{t-1} \sim N(0, h_t^2)$$

which implies:

$$\kappa(\eta_t | \Omega_{t-1}) = \frac{E[\eta_t^4 | \Omega_{t-1}]}{E[\eta_t^2 | \Omega_{t-1}]^2} = 3.$$

By inverting this relation we have a conditional 4th moment:

$$E[\eta_t^4 | I_{t-1}] = 3E[\eta_t^2 | I_{t-1}]^2 = 3h_t^4$$

and the unconditional 4th moment:

$$m_4 = E[\eta_t^4] = E[E[\eta_t^4 | I_{t-1}]] = E[3h_t^4] = 3E[h_t^4].$$

Moreover we have that

$$h_t^4 = (\alpha_0 + \alpha_1 \eta_{t-1}^2 + \beta_1 h_{t-1}^2)^2$$

Using this we have:

$$\begin{aligned} E[\eta_t^4] &= 3E[h_t^4] = 3E[(\alpha_0 + \alpha_1 \eta_{t-1}^2 + \beta_1 h_{t-1}^2)^2] \\ &= 3E[\alpha_0^2 + \alpha_1^2 \eta_{t-1}^4 + \beta_1^2 h_{t-1}^4 + 2\alpha_0 \alpha_1 \eta_{t-1}^2 + 2\alpha_0 \beta_1 h_{t-1}^2 + 2\alpha_1 \eta_{t-1}^2 \beta_1 h_{t-1}^2] \\ &= 3\alpha_0^2 + 3\alpha_1^2 E[\eta_{t-1}^4] + 3\beta_1^2 E[h_{t-1}^4] + 6\alpha_0 \alpha_1 E[\eta_{t-1}^2] + 6\alpha_0 \beta_1 E[h_{t-1}^2] + 6\alpha_1 \beta_1 E[\eta_{t-1}^2 h_{t-1}^2] \\ &= 3\alpha_0^2 + 3\alpha_1^2 E[\eta_{t-1}^4] + 3\beta_1^2 E[h_{t-1}^4] + 6\alpha_0 \alpha_1 E[\eta_{t-1}^2] + 6\alpha_0 \beta_1 E[h_{t-1}^2] + 6\alpha_1 \beta_1 E[h_{t-1}^4] \end{aligned}$$

because of stationarity we have that

$$\begin{aligned} m_4 &= E[\eta_t^4] = E[\eta_{t-1}^4] = 3E[h_t^4] = 3E[h_{t-1}^4] \\ m_2 &= E[\eta_{t-1}^2] = E[h_{t-1}^2] = \sigma^2 = \alpha_0/(1 - \alpha_1 - \beta_1) \end{aligned}$$

so we can write

$$m_4 = 3\alpha_0^2 + 3\alpha_1^2 m_4 + \beta_1^2 m_4 + 6\alpha_0 \alpha_1 \sigma^2 + 6\alpha_0 \beta_1 \sigma^2 + 2\alpha_1 \beta_1 m_4$$

solving for m_4

$$\begin{aligned} m_4 - 3\alpha_1^2 m_4 - \beta_1^2 m_4 - 2\alpha_1\beta_1 m_4 &= 3\alpha_0^2 + 6\alpha_0\alpha_1\sigma^2 + 6\alpha_0\beta_1\sigma^2 \\ m_4(1 - 3\alpha_1^2 - \beta_1^2 - 2\alpha_1\beta_1) &= 3\alpha_0^2 + 6\alpha_0\alpha_1\sigma^2 + 6\alpha_0\beta_1\sigma^2 \\ &= \alpha_0(3\alpha_0 + 6\alpha_1\sigma^2 + 6\beta_1\sigma^2) \\ &= \sigma^2(1-\alpha_1-\beta_1)(3\sigma^2(1-\alpha_1-\beta_1) + 6\alpha_1\sigma^2 + 6\beta_1\sigma^2) \\ &= 3\sigma^4(1-\alpha_1-\beta_1)((1-\alpha_1-\beta_1) + 2\alpha_1 + 2\beta_1) \\ &= 3\sigma^4(1-\alpha_1-\beta_1)(1+\alpha_1+\beta_1) \\ &= 3\sigma^4(1-\{\alpha_1+\beta_1\})(1+\{\alpha_1+\beta_1\}) \\ &= 3\sigma^4(1-\{\alpha_1+\beta_1\}^2) \end{aligned}$$

so we have

$$m_4 = \frac{3\sigma^4(1-\{\alpha_1+\beta_1\}^2)}{1-3\alpha_1^2-\beta_1^2-2\alpha_1\beta_1} = \frac{3\sigma^4(1-\{\alpha_1+\beta_1\}^2)}{1-2\alpha_1^2-\{\alpha_1+\beta_1\}^2}$$

For m_4 to exists it must be that $1-2\alpha_1^2-(\alpha_1+\beta_1)^2 > 0$.

5. The kurtosis is:

$$\frac{m_4}{m_2^2} = \frac{3\sigma^4[1-(\alpha_1+\beta_1)^2]}{\sigma^4(1-2\alpha_1^2-(\alpha_1+\beta_1)^2)} = \left(\frac{1-(\alpha_1+\beta_1)^2}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2}\right) \times 3 \geq 3$$

For the Kurtosis to exists it must be that $1-(\alpha_1+\beta_1)^2-2\alpha_1^2 > 0$. If that is the case then

$$1-(\alpha_1+\beta_1)^2 > 1-2\alpha_1^2-(\alpha_1+\beta_1)^2$$

which is always true (unless $\alpha_1 = 0$). The excess kurtosis is

$$\begin{aligned} \frac{m_4}{m_2^2} - 3 &= \left(\frac{1-(\alpha_1+\beta_1)^2}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2}\right) \times 3 - 3 \\ &= \left(\frac{1-(\alpha_1+\beta_1)^2}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2}\right) \times 3 - 3 \\ &= \left(\frac{1-(\alpha_1+\beta_1)^2}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2} - 1\right) \times 3 \\ &= \frac{1-(\alpha_1+\beta_1)^2-(1-(\alpha_1+\beta_1)^2-2\alpha_1^2)}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2} \times 3 \\ &= \frac{6\alpha_1^2}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2}. \end{aligned}$$

Solution to question 4

1. There are the following cases:

	$\alpha \neq 0$		$\alpha = 0$	
	$\lvert\theta\rvert < 1$	$\theta = 1$	$\lvert\theta\rvert < 1$	$\theta = 1$
$\lvert\rho\rvert < 1$	A	B	E	F
$\rho = 1$	C	D	G	H

   Starting with the cases with $\alpha = 0$. In all of these cases, there cannot be cointegration. The degree of integration of the series will be as follows. In case E both I(0), case F $y_t \sim I(0), x_t \sim I(1)$, case G $x_t \sim I(1), y_t \sim I(0)$, case H both series are I(1), but not cointegrated.

   Considering now the cases with $\alpha \neq 0$. In case A both series are stationary. In case B $x_t \sim I(1)$ and $y_t \sim I(1)$ and y_t shares the same stochastic trend of x_t thanks to the α. This is the cointegration case. In case C we have that $x_t \sim I(0)$ while $y_t \sim I(1)$ for every value of α. The variables cannot be cointegrared. Finally in case D both series are I(1) but they are not cointegrated since each variable has its own stochastic trend.

2. Rewrite the model:

$$\begin{aligned} x_t &= u_{1t} = \theta u_{1t-1} + \epsilon_{1t} = \theta x_{t-1} + \epsilon_{1t} \\ y_t - \alpha x_t &= u_{2t} = \rho u_{2t-1} + \epsilon_{2t} = \rho(y_{t-1} - \alpha x_{t-1}) + \epsilon_{2t} \end{aligned}$$

   The VAR representation is:

$$\begin{aligned} \begin{bmatrix} 1 & 0 \\ -\alpha & 1 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} &= \begin{bmatrix} \theta & 0 \\ -\alpha\rho & \rho \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \\ \begin{bmatrix} x_t \\ y_t \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ -\alpha & 1 \end{bmatrix}^{-1} \begin{bmatrix} \theta & 0 \\ -\alpha\rho & \rho \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ -\alpha & 1 \end{bmatrix}^{-1} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \\ &= \begin{bmatrix} \theta & 0 \\ \alpha(\theta - \rho) & \rho \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \\ &= \begin{bmatrix} \theta x_{t-1} + \epsilon_{1t} \\ \rho y_{t-1} + \alpha(\theta - \rho) x_{t-1} + \alpha\epsilon_{1t} + \epsilon_{2t} \end{bmatrix} \end{aligned}$$

3. The MA is obtained by inverting the AR, which can be done under the assumption

of stationarity. Start from the VAR.

$$\begin{aligned}
\begin{bmatrix} x_t \\ y_t \end{bmatrix} &= \begin{bmatrix} \theta & 0 \\ \alpha(\theta-\rho) & \rho \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \\
\begin{bmatrix} 1-\theta L & 0 \\ -\alpha(\theta-\rho)L & 1-\rho L \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}
\end{aligned}$$

The roots of the characteristic polynomial

$$\left| L^2\theta\rho - L\rho - L\theta + 1 \right| = 0$$

are: $\left\{ \frac{1}{\theta}, \frac{1}{\rho} \right\}$. Hence we can invert this under the assumption that both θ and ρ are in the unit circle.

$$\begin{aligned}
\begin{bmatrix} x_t \\ y_t \end{bmatrix} &= \begin{bmatrix} 1-\theta L & 0 \\ -\alpha(\theta-\rho)L & 1-\rho L \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{1-\theta L} & 0 \\ \frac{1}{1-\theta L}\frac{1}{1-\rho L}\alpha(\theta-\rho)L & \frac{1}{1-\rho L} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{1-\theta L}\epsilon_{1t} \\ \alpha\left(\frac{1}{1-L\rho} + \frac{(\theta-\rho)L}{(1-L\theta)(1-L\rho)}\right)\epsilon_{1t} + \frac{1}{1-\rho L}\epsilon_{2t} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{1-\theta L}\epsilon_{1t} \\ \alpha\frac{1}{1-\theta L}\epsilon_{1t} + \frac{1}{1-\rho L}\epsilon_{2t} \end{bmatrix} \\
&= \begin{bmatrix} \Sigma_{j=1}^{\infty}\theta^j\epsilon_t \\ \alpha\Sigma_{j=1}^{\infty}\theta^j\epsilon_t + \Sigma_{j=1}^{\infty}\rho^j\epsilon_t \end{bmatrix}
\end{aligned}$$

4. Finally, for the VEC we have to assume that there is cointegration, hence we set $|\rho| < 1,\ \theta = 1$ and $\alpha \neq 0$. The VAR is:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \alpha(1-\rho) & \rho \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

subtract on both sides

$$\begin{aligned}
\begin{bmatrix} x_t \\ y_t \end{bmatrix} - \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} &= \begin{bmatrix} 1-1 & 0 \\ \alpha(1-\rho) & \rho-1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 \\ \alpha(1-\rho) & \rho-1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}
\end{aligned}$$

which shows that the autoregressive matrix has reduced rank. This matrix can be obtained as the product of $\begin{bmatrix} 0 \\ (1-\rho) \end{bmatrix}$ with the cointegrating vector:

$$
\begin{aligned}
\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} &= \begin{bmatrix} 0 \\ (1-\rho) \end{bmatrix} \begin{bmatrix} -\alpha & 1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \\
\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} &= \begin{bmatrix} 0 \\ (1-\rho) \end{bmatrix} (y_{t-1} - \alpha x_{t-1}) + \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \\
\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} &= \begin{bmatrix} \epsilon_t \\ (\rho - 1)(\alpha x_{t-1} - y_{t-1}) + \epsilon_{2t} + \alpha \epsilon_t \end{bmatrix}
\end{aligned}
$$

5. Imposing the representations with cointegration pays off in terms of efficiency of the estimates. However these representation imposes specific constraints on the coefficients so if those are not satisfied using these representations will induce misspecification error.