Solutions for Preparation Exam
Introduction to Time Series

Bachelor Econometrics and Operations Research
Faculty of Economics and Business Administration

Exam:	Introduction to Time Series
Code:	E_EOR3_ITS
Coordinator:	dr. F. Blasques
Date:	–
Time:	–
Duration:	2 hours and 45 minutes
Calculator:	Not allowed
Graphical calculator:	Not allowed
Number of questions:	4
Type of questions:	Open
Answer in:	English
Credit score:	100 credits counts for a 10
Grades:	Made public within 10 working days
Inspection:	By appointment (send e-mail to f.blasques@vu.nl)
Number of pages:	3, including front page

- Read the entire exam carefully before you start answering the questions.
- Be clear and concise in your statements, but justify every step in your derivations.
- The questions should be handed back at the end of the exam. Do not take it home.

Good luck!

Question 1 [35 points] ARMA Models

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a time-series generated by an ARMA$(2,2)$ model

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} \quad , \quad t \in \mathbb{Z}\ ,$$

where $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of white noise innovations with variance $\sigma^2_\epsilon > 0$.

(a) (**6 pts**) Give the definitions of strict stationarity and weak stationarity. Can you give an example of a strictly stationary time-series that is not weakly stationary? Justify your answer.

Answer:

[**2pts**] A time series $\{X_t\}_{t\in\mathbb{Z}}$ is strictly stationary if the distribution of any finite subvector is invariant in time $(X_1,\ldots,X_h) \overset{d}{=} (X_t,\ldots,X_{t+h})$ for all t and h.

[**2pts**] A time series is weakly stationary if it has a mean, variance and autocovariances that are finite and constant over time, $\mathbb{E}(X_t) = \mathbb{E}(X_{t+h})$ for all h, $\mathbb{V}\text{ar}(X_t) = \mathbb{V}\text{ar}(X_{t+h})$ for all h, and $\mathbb{C}\text{ov}(X_t, X_{t-h}) = \gamma(h)$ for all t and all h.

[**2pts**] An IID sequence of Cauchy random variables is strictly stationary since it is IID. However, a Cauchy variable doesn't have finite variance. Therefore, a sequence of Cauchy random variables is not weakly stationary.

(b) (**8 pts**) Please rewrite the ARMA$(2,2)$ model in lag polynomial form $\phi(L)X_t = \theta(L)\epsilon_t$. Give an expression for the polynomials $\phi(L)$ and $\theta(L)$. Use the general weak stationarity theorem to show that $\{X_t\}_{t\in\mathbb{Z}}$ is weakly stationary if $\phi(L)$ is invertible.

Answer:

[**2pts**] In lag polynomial notation, the ARMA(2,2) can be written as

$$(1 - \phi_1 L - \phi_2 L^2)X_t = (1 + \theta_1 L + \theta_2 L^2)\epsilon_t$$

or

$$\phi(L)X_t = \theta(L)\epsilon_t$$

[**2pts**] Where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2$ and $\theta(L) = 1 + \theta_1 L + \theta_2 L^2$.

[**4pts**] If $\phi(L)$ is invertible, then we know that $\{X_t\}$ can be written as a weighted average of the white noise sequence $\{\epsilon_t\}$

$$X_t = \phi^{-1}(L)\theta(L)\epsilon_t = \frac{\theta(L)}{\phi(L)}\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}.$$

with absolutely summable coefficients $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Finally, since the innovations $\{\epsilon_t\}$ are white noise, they are also weakly stationary, and we can conclude that $\{X_t\}$ is weakly stationary by the general weak stationarity theorem.

(c) (**11pts**) Suppose that $|\phi_1| < 1$, $\theta_1 \neq 0$ and $\phi_2 = \theta_2 = 0$. Calculate the unconditional mean and variance of $\{X_t\}_{t\in\mathbb{Z}}$. In other words, derive $\mathbb{E}(X_t)$ and $\mathbb{V}\text{ar}(X_t)$.

Answer:

[**2pts**] Given the parameter restrictions, we have an ARMA(1,1) model

$$X_t = \phi_1 X_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}$$

$$\Leftrightarrow \qquad \phi(L)X_t = \theta(L)\epsilon_t$$

Furthermore, since $|\phi_1| < 1$, the autoregressive lag polynomial is invertible, and hence $\{X_t\}$ is a weighted average of a white noise sequence

$$X_t = \phi^{-1}(L)\theta(L)\epsilon_t = \frac{\theta(L)}{\phi(L)}\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}.$$

with absolutely summable coefficients $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Since the innovations $\{\epsilon_t\}$ are white noise, they are also weakly stationary, and we conclude that $\{X_t\}$ is weakly stationary by the general weak stationarity theorem.

[**4pts**] Therefore we can calculate the unconditional mean as follows

$$\begin{aligned}
\mathbb{E}(X_t) &= \mathbb{E}(\phi_1 X_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t) \\
&= \phi_1\,\mathbb{E}(X_{t-1}) + \theta_1\,\mathbb{E}(\epsilon_{t-1}) + \mathbb{E}(\epsilon_t) \qquad \text{(by linearity of expectation)} \\
&= \phi_1\,\mathbb{E}(X_{t-1}) \qquad (\mathbb{E}(\epsilon_t) = \mathbb{E}(\epsilon_{t-1}) = 0 \text{ because } \{\epsilon_t\} \text{ is white noise}) \\
&= \phi_1\,\mathbb{E}(X_t) \qquad (\mathbb{E}(X_t) = \mathbb{E}(X_{t-1}) \text{ because } \{X_t\} \text{ is weakly stationary})
\end{aligned}$$

Finally, since $|\phi_1| < 1$, the mean can only be $\mathbb{E}(X_t) = 0$.

[**5pts**] The unconditional variance is given by

$$\begin{aligned}
\mathbb{V}\text{ar}(X_t) &= \mathbb{V}\text{ar}(\phi_1 X_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t) \\
&= \phi_1^2 \mathbb{V}\text{ar}(X_{t-1}) + \theta_1^2\,\mathbb{V}\text{ar}(\epsilon_{t-1}) + \mathbb{V}\text{ar}(\epsilon_t) \\
&\quad + 2\phi_1\theta_1 \mathbb{C}\text{ov}(X_{t-1}, \epsilon_{t-1}) + 2\phi_1 \mathbb{C}\text{ov}(X_{t-1}, \epsilon_t)
\end{aligned}$$

We now note that:

(i) $\mathbb{V}\text{ar}(X_{t-1}) = \mathbb{V}\text{ar}(X_t)$ because $\{X_t\}$ is weakly stationary

(ii) $\mathbb{V}\text{ar}(\epsilon_{t-1}) = \mathbb{V}\text{ar}(\epsilon_t) = \sigma_\epsilon^2$ because $\{\epsilon_t\}$ is white noise with variance σ_ϵ^2

(iii) $\mathbb{C}\text{ov}(X_{t-1}, \epsilon_t) = 0$ because $\{X_t\}$ is generated by an ARMA model with invertible AR polynomial. This implies that X_t can be written as a weighted average of past innovations $X_{t-1} = \sum_{j=1}^{\infty} \psi_j \epsilon_{t-j}$. Hence

$$\mathbb{C}\text{ov}(X_{t-1}, \epsilon_t) = \mathbb{C}\text{ov}(\sum_{j=1}^{\infty} \psi_j \epsilon_{t-j}, \epsilon_t) = \sum_{j=1}^{\infty} \psi_j \mathbb{C}\text{ov}(\epsilon_{t-j}, \epsilon_t) = 0.$$

All the covariances $\mathbb{C}\text{ov}(\epsilon_{t-j}, \epsilon_t)$ in the sum above are zero because $\{\epsilon_t\}$ is white noise (and hence, the innovations are uncorrelated).

(iv) $\mathbb{C}\text{ov}(X_{t-1}, \epsilon_{t-1}) = \sigma^2$ because

$$\begin{aligned}
\mathbb{C}\text{ov}(X_{t-1}, \epsilon_{t-1}) &= \mathbb{C}\text{ov}(\phi_1 X_{t-2} + \theta_1 \epsilon_{t-2} + \epsilon_{t-1}, \epsilon_{t-1}) \\
&\quad \text{(by definition of } X_{t-1}) \\
&= \phi_1 \mathbb{C}\text{ov}(X_{t-2}, \epsilon_{t-1}) + \theta_1 \mathbb{C}\text{ov}(\epsilon_{t-2}, \epsilon_{t-1}) + \mathbb{C}\text{ov}(\epsilon_{t-1}, \epsilon_{t-1}) \\
&\quad (\mathbb{C}\text{ov}(\epsilon_{t-2}, \epsilon_{t-1}) = 0 \text{ by the same argument as in point (iii) above)} \\
&\quad (\mathbb{C}\text{ov}(\epsilon_{t-2}, \epsilon_{t-1}) = 0 \text{ because } \{\epsilon_t\} \text{ is white noise [uncorrelated])} \\
&= \mathbb{V}\text{ar}(\epsilon_{t-1}) = \sigma_\epsilon^2. \\
&\quad (\text{because } \mathbb{C}\text{ov}(\epsilon_{t-1}, \epsilon_{t-1}) = \mathbb{V}\text{ar}(\epsilon_{t-1}) = \sigma_\epsilon^2)
\end{aligned}$$

Taking into account the points (i), (ii), (iii) and (iv) above, we can thus simplify the expression for the variance of X_t as follows:

$$\mathbb{V}\text{ar}(X_t) = \phi_1^2 \mathbb{V}\text{ar}(X_t) + (2\phi_1\theta_1 + \theta_1^2 + 1)\sigma_\epsilon^2.$$

Finally, solving for the variance yields,

$$\mathbb{V}\text{ar}(X_t) = \frac{(2\phi_1\theta_1 + \theta_1^2 + 1)\sigma_\epsilon^2}{(1 - \phi_1^2)}.$$

(d) (**10pts**) Suppose now that $\phi_1 \neq 0$ and $\phi_2 \neq 0$ and $\theta_1 = \theta_2 = 0$. Additionally, assume that the innovations $\{\epsilon_t\}_{t\in\mathbb{Z}}$ are independent and identically distributed (iid) Gaussian random variables $\{\epsilon_t\}_{t\in\mathbb{Z}} \sim \text{NID}(0, \sigma_\epsilon^2)$. Produce the 2-step ahead forecast and the variance of the 2-step ahead forecast error. Derive 95% confidence bounds for your forecast.

Answer:

[**3pts**] Given the parameter restrictions, we obtain an AR(2) model. The 2-step-ahead forecast is given by

$$\begin{aligned}
\hat{X}_{T+2} &= \mathbb{E}(X_{T+2}|X_1, ..., X_T) \\
&= \mathbb{E}(\phi_1 X_{T+1} + \phi_2 X_T + \epsilon_{T+2}|X_1, ..., X_T) \\
&\quad \text{(by the AR(2) definition of } X_{T+2}) \\
&= \phi_1 \mathbb{E}(X_{T+1}|X_1, ..., X_T) + \phi_2 \mathbb{E}(X_T|X_1, ..., X_T) + \mathbb{E}(\epsilon_{T+2}|X_1, ..., X_T) \\
&\quad \text{(by linearity of expectation)} \\
&= \phi_1 \hat{X}_{T+1} + \phi_2 X_T + \mathbb{E}(\epsilon_{T+2}) \\
&\quad (\text{because } \hat{X}_{T+1} = \mathbb{E}(X_{T+1}|X_1, ..., X_T) \text{ by definition, } X_T \text{ is given, and } \epsilon_{T+2} \text{ is} \\
&\quad\quad \text{independent of } X_1, ..., X_T \text{ since } \{X_t\} \text{ is generated by an ARMA model)} \\
&= \phi_1 \hat{X}_{T+1} + \phi_2 X_T \\
&\quad (\mathbb{E}(\epsilon_{T+2}) = 0 \text{ since } \{\epsilon_t\} \text{ iid with mean zero)}
\end{aligned}$$

The 1-step-ahead forecast is given by

$$\begin{aligned}
\hat{X}_{T+1} &= \mathbb{E}(X_{T+1}|X_1, ..., X_T) \\
&= \mathbb{E}(\phi_1 X_T + \phi_2 X_{T-1} + \epsilon_{T+1}|X_1, ..., X_T) \\
&\quad \text{(by the AR(2) definition of } X_{T+1}) \\
&= \phi_1 \mathbb{E}(X_T|X_1, ..., X_T) + \phi_2 \mathbb{E}(X_{T-1}|X_1, ..., X_T) + \mathbb{E}(\epsilon_{T+1}|X_1, ..., X_T) \\
&\quad \text{(by linearity of expectation)} \\
&= \phi_1 X_T + \phi_2 X_{T-1} + \mathbb{E}(\epsilon_{T+1}) \\
&\quad \text{(because } X_T \text{ and } X_{T-1} \text{ are known, and } \epsilon_{T+1} \text{ is} \\
&\qquad \text{independent of } X_1, ..., X_T \text{ since } \{X_t\} \text{ is generated by an ARMA model)} \\
&= \phi_1 X_T + \phi_2 X_{T-1} \\
&\quad (\mathbb{E}(\epsilon_{T+1}) = 0 \text{ since } \{\epsilon_t\} \text{ is iid with mean zero)}
\end{aligned}$$

Hence 2-step-ahead forecast simplifies to

$$\hat{X}_{T+2} = \phi_1 \hat{X}_{T+1} + \phi_2 X_T = (\phi_1^2 + \phi_2) X_T + \phi_1 \phi_2 X_{T-1}.$$

[**3pts**] The 2-step-ahead forecast error is given by

$$\begin{aligned}
e_{T+2} &= X_{T+2} - \hat{X}_{T+2} \\
&= \phi_1 X_{T+1} + \phi_2 X_T + \epsilon_{T+2} - \phi_1 \hat{X}_{T+1} - \phi_2 X_T \\
&= \phi_1 X_{T+1} - \phi_1 \hat{X}_{T+1} + \epsilon_{T+2} \\
&= \phi_1 (X_{T+1} - \hat{X}_{T+1}) + \epsilon_{T+2}
\end{aligned}$$

The 1-step-ahead forecast error is given by

$$\begin{aligned}
e_{T+1} &= X_{T+1} - \hat{X}_{T+1} \\
&= \phi_1 X_T + \phi_2 X_{T-1} + \epsilon_{T+1} - \phi_1 X_T - \phi_2 X_{T-1} \\
&= \epsilon_{T+1}
\end{aligned}$$

Hence, we the 2-step-ahead forecast error simplifies to

$$e_{T+2} = \phi_1 (X_{T+1} - \hat{X}_{T+1}) + \epsilon_{T+2} = \phi_1 \epsilon_{T+1} + \epsilon_{T+2}$$

[**2pts**] The variance of the 2-step-ahead forecast is thus given by

$$\mathbb{V}\text{ar}(e_{T+2}) = \phi_1^2 \mathbb{V}\text{ar}(\epsilon_{T+1}) + \mathbb{V}\text{ar}(\epsilon_{T+2}) = (\phi_1^2 + 1)\sigma_\epsilon^2$$

because the innovations $\{\epsilon_t\}$ are independent with constant variance σ_ϵ^2.

[**2pts**] Finally, if the innovations are Gaussian, then the 95% confidence bounds are then given by $\hat{X}_{T+2} \pm 1.96\sqrt{\mathbb{V}\text{ar}(e_{T+2})}$, and can be written as

$$\hat{X}_{T+2} \pm 1.96\sqrt{(\phi_1^2 + 1)\sigma_\epsilon^2}.$$

Question 2 [15 points] ML Estimation

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a time-series generated by a MA(1) process,

$$X_t = \theta_1 \epsilon_{t-1} + \epsilon_t \quad , \quad t \in \mathbb{Z} \ ,$$

with independent and identical Gaussian innovations $\{\epsilon_t\}_{t\in\mathbb{Z}} \sim \text{NID}(0, \sigma_\epsilon^2)$ and $\sigma_\epsilon^2 > 0$. Note that this implies that ϵ_t has the following probability density function:

$$f(\epsilon_t; \sigma_\epsilon^2) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} e^{-\epsilon_t^2/2\sigma_\epsilon^2} \quad , \quad t \in \mathbb{Z}$$

(a) (**8pts**) Give an expression of the log likelihood function for the unknown parameters $(\theta_1, \sigma_\epsilon^2)$ using the joint Gaussianity of the sample $X_1, ..., X_T$.

Answer:

[**2pts**] If $\mathbf{X}_T := (X_1, ..., X_T)$ is jointly Gaussian, then the log likelihood for the vector of parameters $\psi := (\theta_1, \sigma_\epsilon^2)$ is given by,

$$\hat{\psi}_T = \arg\max_{\psi} -\frac{T}{2}\log 2\pi - \frac{1}{2}\log|\Gamma(\psi)| - \frac{1}{2}(\mathbf{X}_T' \Gamma^{-1}(\psi)\mathbf{X}_T).$$

[**2pts**] Furthermore, the variance-covariance matrix $\Gamma^{-1}(\psi)$ is given by

$$\Gamma(\psi) = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \dots & \gamma(T) \\ \gamma(1) & \gamma(0) & \gamma(1) & \dots & \gamma(T-1) \\ \gamma(2) & \gamma(1) & \gamma(0) & \dots & \gamma(T-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(T) & \gamma(T-1) & \gamma(T-2) & \dots & \gamma(0) \end{bmatrix}.$$

[**4pts**] where the variance is given by

$$\begin{aligned} \gamma(0) = \mathbb{V}\text{ar}(X_t) &= \mathbb{V}\text{ar}(\theta_1\epsilon_{t-1} + \epsilon_t) \\ &\quad \text{(by MA(1) definition of } X_t) \\ &= \theta_1 \mathbb{V}\text{ar}(\epsilon_{t-1}) + \mathbb{V}\text{ar}(\epsilon_t) \\ &\quad (\text{because } \epsilon_t \perp \epsilon_{t-1} \text{ since } \{\epsilon_t\} \sim\text{NID}(0, \sigma_\epsilon^2)) \\ &= \theta_1\sigma_\epsilon^2 + \sigma_\epsilon^2 \\ &\quad (\mathbb{V}\text{ar}(\epsilon_t) = \sigma_\epsilon^2 \ \forall\, t \text{ since } \{\epsilon_t\} \sim\text{NID}(0, \sigma_\epsilon^2)) \\ &= (1 + \theta_1^2)\sigma_\epsilon^2 \end{aligned}$$

and the first-order autocovariance is given by

$$\begin{aligned}
\gamma(1) = \mathbb{C}\text{ov}(X_t, X_{t-1}) &= \mathbb{C}\text{ov}(\theta_1\epsilon_{t-1} + \epsilon_t, \theta_1\epsilon_{t-2} + \epsilon_{t-1}) \\
&\quad \text{(by MA(1) definition of } X_t \text{ and } X_{t-1}) \\
&= \theta_1^2\mathbb{C}\text{ov}(\epsilon_{t-1}, \epsilon_{t-2}) + \theta_1\mathbb{C}\text{ov}(\epsilon_{t-1}, \epsilon_{t-1}) \\
&\quad + \theta_1\mathbb{C}\text{ov}(\epsilon_t, \epsilon_{t-2}) + \mathbb{C}\text{ov}(\epsilon_t, \epsilon_{t-1}) \\
&\quad \text{(by linearity of the covariance)} \\
&= \theta_1\mathbb{C}\text{ov}(\epsilon_{t-1}, \epsilon_{t-1}) \\
&\quad (\mathbb{C}\text{ov}(\epsilon_t, \epsilon_s) = 0 \ \forall\ t \neq s \text{ because } \{\epsilon_t\} \sim \text{NID}(0, \sigma_\epsilon^2)) \\
&= \theta_1\mathbb{V}\text{ar}(\epsilon_{t-1}) \\
&\quad (\text{because } \mathbb{C}\text{ov}(\epsilon_{t-1}, \epsilon_{t-1}) = \mathbb{V}\text{ar}\epsilon_{t-1}) \\
&= \theta_1\sigma_\epsilon^2. \\
&\quad (\mathbb{V}\text{ar}(\epsilon_{t-1}) = \sigma_\epsilon^2 \text{ since } \{\epsilon_t\} \sim \text{NID}(0, \sigma_\epsilon^2))
\end{aligned}$$

Finally, for any lag $h \geq 2$ we have

$$\begin{aligned}
\gamma(h) = \mathbb{C}\text{ov}(X_t, X_{t-h}) &= \mathbb{C}\text{ov}(\theta_1\epsilon_{t-1} + \epsilon_t, \theta_1\epsilon_{t-h-1} + \epsilon_{t-h}) \\
&\quad \text{(by MA(1) definition of } X_t \text{ and } X_{t-h}) \\
&= \theta_1^2\mathbb{C}\text{ov}(\epsilon_{t-1}, \epsilon_{t-h-1}) + \theta_1\mathbb{C}\text{ov}(\epsilon_{t-1}, \epsilon_{t-h}) \\
&\quad + \theta_1\mathbb{C}\text{ov}(\epsilon_t, \epsilon_{t-h-1}) + \mathbb{C}\text{ov}(\epsilon_t, \epsilon_{t-h}) \\
&\quad \text{(by linearity of the covariance)} \\
&= 0 \\
&\quad (\mathbb{C}\text{ov}(\epsilon_t, \epsilon_s) = 0 \ \forall\ t \neq s \text{ as innovations } \{\epsilon_t\} \text{ are independent})
\end{aligned}$$

We thus have that

$$\Gamma(\theta_1, \sigma_\epsilon^2) = \begin{bmatrix} (1+\theta_1^2)\sigma_\epsilon^2 & \theta_1\sigma_\epsilon^2 & 0 & \dots & 0 \\ \theta_1\sigma_\epsilon^2 & (1+\theta_1^2)\sigma_\epsilon^2 & \theta_1\sigma_\epsilon^2 & \dots & 0 \\ 0 & \theta_1\sigma_\epsilon^2 & (1+\theta_1^2)\sigma_\epsilon^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & (1+\theta_1^2)\sigma_\epsilon^2 \end{bmatrix}.$$

(b) (**7pts**) Write down the conditional likelihood function using prediction error decomposition starting at $t = 2$.

Answer:

[**2pts**] First, we note that the sequence $\epsilon_1, ..., \epsilon_t$ is known conditional on observing $x_1, ..., x_t$.

[**2pts**] Hence, given our MA(1) model, we have

$$\begin{aligned}
X_2|X_1 = X_2|\epsilon_1 &\sim N(\theta_1\epsilon_1, \sigma_\epsilon^2) \\
X_3|X_2, X_1 = X_3|\epsilon_2, \epsilon_1 &\sim N(\theta_1\epsilon_2, \sigma_\epsilon^2)
\end{aligned}$$

and in general

$$X_t|X_{t-1}, X_{t-2}, ... = X_t|\epsilon_{t-1}, \epsilon_{t-2}, ... \sim N(\theta_1\epsilon_{t-1}, \sigma_\epsilon^2)$$

[**3pts**] Second, we note that the joint density of the data $f(x_1, ..., x_T, \theta_1, \sigma_\epsilon^2)$ can be factorized as a product of conditional densities (ignoring the first marginal)

$$
\begin{aligned}
f(x_1, ..., x_T; \theta, \sigma^2) &\approx \prod_{t=2}^{T} f(x_t | x_{t-1}, ..., x_1; \theta_1, \sigma^2) \\
&= \prod_{t=2}^{T} f(x_t | \epsilon_{t-1}, ..., \epsilon_1; \theta_1, \sigma^2) \\
&= \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big[- \frac{(x_t - \theta_1 \epsilon_{t-1})^2}{2\sigma_\epsilon^2} \Big].
\end{aligned}
$$

Question 3 [15 points]   Unit-Root Testing

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a time-series generated by an AR(2) process,

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t \quad , \quad t \in \mathbb{Z} \ ,$$

where $\{\epsilon_t\}_{t\in\mathbb{Z}} \sim \text{WN}(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon^2 > 0$ and consider the ADF regression,

$$\Delta X_t = \beta X_{t-1} - \phi_2 \Delta X_{t-1} + \epsilon_t.$$

(a) (**7pts**) Re-write the AR(2) model in the ADF regression form. Show that testing the hypothesis $\text{H}_0 : \beta = 0$ is equivalent to testing for a unit root in the polynomial $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$.

Answer:

[**4pts**] Re-writing the AR(2) in ADF form yields:

$$\begin{aligned}
& X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t \\
\Leftrightarrow \quad & X_t - X_{t-1} = \phi_1 X_{t-1} - X_{t-1} + \phi_2 X_{t-2} + \epsilon_t \\
& \text{(subtracting } X_{t-1} \text{ on both sides)} \\
\Leftrightarrow \quad & \Delta X_t = (\phi_1 - 1) X_{t-1} + \phi_2 X_{t-2} + \epsilon_t \\
& \text{(noting that } X_t - X_{t-1} = \Delta X_t) \\
\Leftrightarrow \quad & \Delta X_t = (\phi_1 - 1) X_{t-1} + \phi_2 X_{t-1} - \phi_2 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t \\
& \text{(adding and subtracting } \phi_2 X_{t-1} \text{ on the right-hand-side)} \\
\Leftrightarrow \quad & \Delta X_t = (\phi_1 + \phi_2 - 1) X_{t-1} - \phi_2 \Delta X_{t-1} + \epsilon_t \\
& \text{(noting that } -\phi_2 X_{t-1} + \phi_2 X_{t-2} = -\phi_2 \Delta X_{t-1}) \\
\Leftrightarrow \quad & \Delta X_t = \beta X_{t-1} - \phi_2 \Delta X_{t-1} + \epsilon_t \\
& \text{(defining } \beta = (\phi_1 + \phi_2 + 1))
\end{aligned}$$

[**2pts**] A unit root occurs when $\phi(1) = 0$. Since our polynomial $\phi(L)$ takes the form $\phi(L) = 1 - \phi_1 L - \phi_2 L^2$, the unit root hypothesis states that

$$\phi(1) = 1 - \phi_1 \times 1 - \phi_2 \times 1 = 1 - \phi_1 - \phi_2 = 0$$

[**1pts**] As a result, testing if $\phi(1) = 0$ is equivalent to testing if $1 - \phi_1 - \phi_2 = 0$, which, in turn, is equivalent to testing if $\beta = \phi_1 + \phi_2 - 1 = 0$.

[**8pts**] Why is it important to use a general-to-specific approach in the specification of the ADF regression? Justify your answer. **Answer:**

[**2pts**] The ADF test relies on the t-statistic $\hat{\beta}/\text{SE}(\hat{\beta})$ having a Dickey-Fuller distribution under the null hypothesis that $\beta = 0$. However, this t-statistic only has a Dickey-Fuller distribution if the residuals of the ADF regression are white noise.

[**2pts**] When the ADF regression has too few lags, then the residuals will contain autocorrelation. As a result, they will no longer be white noise and the t-statistic will not have the correct distribution.

[**2pts**] On the other hand, when too many lags are included, then estimation uncertainty will be unnecessarily large, and the test results will be less reliable.

[**2pts**] The idea of the general-to-specific approach is to first start with a large number of lags in the ADF regression (enough lags to ensure that the residuals are white noise), and then reduce the ADF regression to the smallest possible number of lags for which the residuals are still white noise. By doing this, we ensure that the t-statistic has the correct distribution and that the estimation uncertainty is as small as possible.

Question 4 [35 points] ADL, Error Correction and Cointegration

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a time-series generated by an ADL$(1,1)$ process,

$$Y_t = \alpha + \phi Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t \quad , \quad t \in \mathbb{Z} \ ,$$

where $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of iid white noise innovations with variance $\sigma_\epsilon^2 > 0$.

(a) (**8pts**) Suppose $|\phi| < 1$. Use the ADL(1,1) model to derive the short-run and long-run multipliers and explain their meaning.

Answer:

[**1pts**] Short-run multiplier is β_0.

[**2pts**] The short-run multiplier measures the expected change in Y_t given a contemporaneous unit increase in X_t.

[**3pts**] The long-run multiplier can be obtained by setting the innovations to zero, $\epsilon_t = 0 \ \forall \ t$, fixing the time series $\{X_t\}$ to some value $\bar{X}$, and solving the ADL equation for a fixed value $\bar{Y}$,

$$\begin{aligned} \bar{Y} &= \alpha + \phi\bar{Y} + \beta_0\bar{X} + \beta_1\bar{X} + 0 \\ \Leftrightarrow \quad \bar{Y}(1-\phi) &= \alpha + (\beta_0 + \beta_1)\bar{X} \\ \Leftrightarrow \quad \bar{Y} &= \frac{\alpha}{1-\phi} + \frac{\beta_0+\beta_1}{1-\phi}\bar{X}. \end{aligned}$$

The long-run multiplier is $\frac{\beta_0+\beta_1}{1-\phi}$.

[**2pts**] The long-run multiplier measures expected long-run change in Y_t given a permanent unit increase in X_t.

(b) (**12pts**) Suppose that $\{X_t\}_{t\in\mathbb{Z}}$ is generated by the following AR(2) model with intercept

$$X_t = \gamma_0 + \gamma_1 X_{t-1} + \gamma_2 X_{t-2} + u_t \quad , \quad t \in \mathbb{Z} \ ,$$

where $\{u_t\}_{t\in\mathbb{Z}}$ is a sequence of white noise iid innovations with variance $\sigma_u^2 > 0$.

Calculate the the impulse response function (IRF) of $\{Y_t\}_{t\in\mathbb{Z}}$ given the origin x for the time-series $\{X_t\}_{t\in\mathbb{Z}}$, the origin y for the time-series $\{Y_t\}_{t\in\mathbb{Z}}$, and a shock of magnitude v in the innovation $\{u_t\}_{t\in\mathbb{Z}}$ at time $t = s$. In particular, give the IRF for $t = s-1$, $t = s$, $t = s+1$ and $t = s+2$.

Answer:

[**2pts**] We have by definition

$$\begin{aligned} \tilde{X}_{s-1} &= x & \qquad \tilde{Y}_{s-1} &= y \\ \tilde{X}_s &= x + v & \qquad \tilde{Y}_s &= y + \beta_0(x+v) \end{aligned}$$

[**5pts**] Furthermore, for periods $s+1$ and $s+2$, we have

$$\begin{aligned} \tilde{X}_{s+1} &= \mathbb{E}(X_{s+1}|\tilde{X}_s, \tilde{X}_{s-1}) \\ &= \mathbb{E}(\gamma_0 + \gamma_1 X_s + \gamma_2 X_{s-1} + \epsilon_{s+1}|\tilde{X}_s, \tilde{X}_{s-1}) \\ &\quad \text{(by definition of } X_{s+1}\text{)} \\ &= \gamma_0 + \gamma_1\, \mathbb{E}(X_s|\tilde{X}_s, \tilde{X}_{s-1}) + \gamma_2\, \mathbb{E}(X_{s-1}|\tilde{X}_s, \tilde{X}_{s-1}) + \mathbb{E}(\epsilon_{s+1}|\tilde{X}_s, \tilde{X}_{s-1}) \\ &\quad \text{(by linearity of the conditional expectation)} \\ &= \gamma_0 + \gamma_1 \tilde{X}_s + \gamma_2 \tilde{X}_{s-1} + \mathbb{E}(\epsilon_{s+1}|\tilde{X}_s, \tilde{X}_{s-1}) \\ &\quad \text{(since } \tilde{X}_s \text{ and } \tilde{X}_{s-1} \text{ are known)} \\ &= \gamma_0 + \gamma_1 \tilde{X}_s + \gamma_2 \tilde{X}_{s-1} + \mathbb{E}(\epsilon_{s+1}) \\ &\quad \text{(since } \epsilon_{s+1} \text{ is independent of past data } X_s, X_{s-1}, ... \\ &\qquad \text{because } \{X_t\} \text{ is generated by an ARMA model)} \\ &= \gamma_0 + \gamma_1 \tilde{X}_s + \gamma_2 \tilde{X}_{s-1} \\ &\quad (\mathbb{E}(\epsilon_t) = 0 \text{ because } \{\epsilon_t\} \text{ is white noise)} \\ &= \gamma_0 + \gamma_1(x+v) + \gamma_2 x \end{aligned}$$

$$
\begin{aligned}
\tilde{X}_{s+2} &= \mathbb{E}(X_{s+2}|\tilde{X}_s, \tilde{X}_{s-1}) \\
&= \mathbb{E}(\gamma_0 + \gamma_1 X_{s+1} + \gamma_2 X_s + \epsilon_{s+2}|\tilde{X}_s, \tilde{X}_{s-1}) \\
&\quad \text{(by definition of } X_{s+2}\text{)} \\
&= \gamma_0 + \gamma_1 \mathbb{E}(X_{s+1}|\tilde{X}_s, \tilde{X}_{s-1}) + \gamma_2 \tilde{X}_s + \mathbb{E}(\epsilon_{s+2}|\tilde{X}_s, \tilde{X}_{s-1}) \\
&\quad \text{(by linearity of the conditional expectation, and because } \tilde{X}_s \text{ is given)} \\
&= \gamma_0 + \gamma_1 \mathbb{E}(X_{s+1}|\tilde{X}_s, \tilde{X}_{s-1}) + \gamma_2 \tilde{X}_s + \mathbb{E}(\epsilon_{s+2}) \\
&\quad \text{(since } \epsilon_{s+2} \text{ is independent of past data } X_s, X_{s-1}, \ldots \\
&\qquad \text{because } \{X_t\} \text{ is generated by an ARMA model)} \\
&= \gamma_0 + \gamma_1 \mathbb{E}(X_{s+1}|\tilde{X}_s, \tilde{X}_{s-1}) + \gamma_2 \tilde{X}_s \\
&\quad (\mathbb{E}(\epsilon_t) = 0 \text{ because } \{\epsilon_t\} \text{ is white noise)} \\
&= (1+\gamma_1)\gamma_0 + (\gamma_1^2 + \gamma_2)\tilde{X}_s + \gamma_1\gamma_2 \tilde{X}_{s-1} \\
&\quad \text{(using the expression for } \tilde{X}_{s+1} = \mathbb{E}(X_{s+1}|\tilde{X}_s, \tilde{X}_{s-1}) \text{ derived above)} \\
&= (1+\gamma_1)\gamma_0 + (\gamma_1^2 + \gamma_2)(x+v) + \gamma_1\gamma_2 x \\
&\quad \text{(using the fact that } \tilde{X}_{s-1} = x\text{)}
\end{aligned}
$$

[**5pts**] For convenience, define $\tilde{D}_s = (\tilde{Y}_s, \tilde{Y}_{s-1}, \tilde{X}_s, \tilde{X}_{s-1})$. Then we finally have,

$$
\begin{aligned}
\tilde{Y}_{s+1} &= \mathbb{E}(Y_{s+1}|\tilde{D}_s) = \mathbb{E}(\alpha + \phi Y_s + \beta_0 X_{s+1} + \beta_1 X_s + \epsilon_{s+1}|\tilde{D}_s) \\
&\quad \text{(by definition of } Y_{s+1}\text{)} \\
&= \alpha + \phi E(Y_s|\tilde{D}_s) + \beta_0 E(X_{s+1}|\tilde{D}_s) + \beta_1 E(X_s|\tilde{D}_s) + \mathbb{E}(\epsilon_{s+1}|\tilde{D}_s) \\
&\quad \text{(by linearity of the conditional expectation)} \\
&= \alpha + \phi E(Y_s|\tilde{D}_s) + \beta_0 E(X_{s+1}|\tilde{D}_s) + \beta_1 E(X_s|\tilde{D}_s) \\
&\quad (\mathbb{E}(\epsilon_{s+1}|\tilde{D}_s) = \mathbb{E}(\epsilon_{s+1}) = 0 \text{ because } \{\epsilon_s\} \text{ is white noise and} \\
&\qquad \{Y_t\} \text{ is generated by an ADL with exogenous regressor } \{X_t\}, \\
&\qquad \text{hence } \epsilon_{s+1} \text{ is independent of } \tilde{D}_s) \\
&= \alpha + \phi \tilde{Y}_s + \beta_0 \tilde{X}_{s+1} + \beta_1 \tilde{X}_s \\
&\quad \text{(since } X_s = \tilde{X}_s \text{ is known and } \tilde{Y}_s \text{ and } \tilde{X}_{s+1} \text{ were derived above)} \\
&= \alpha + \phi(y + \beta_0(x+v)) + \beta_0(\gamma_0 + \gamma_1 \tilde{X}_s + \gamma_2 \tilde{X}_{s-1}) + \beta_1 \tilde{X}_s \\
&\quad \text{(using the expressions of } \tilde{Y}_s = \mathbb{E}(Y_s|\tilde{D}_s) \text{ and } \tilde{X}_{s+1} = \mathbb{E}(X_{s+1}|\tilde{X}_s, \tilde{X}_{s-1}) \text{ derived above)} \\
&= \alpha + \phi(y + \beta_0(x+v)) + \beta_0(\gamma_0 + \gamma_1(x+v) + \gamma_2 x) + \beta_1(x+v) \\
&\quad \text{(since } \tilde{X}_s = x + v \text{ and } \tilde{X}_{s-1} = x\text{)}
\end{aligned}
$$

$$\tilde{Y}_{s+2} = \mathbb{E}(Y_{s+2}|\tilde{D}_s) = \mathbb{E}(\alpha + \phi Y_{s+1} + \beta_0 X_{s+2} + \beta_1 X_{s+1} + \epsilon_{s+2}|\tilde{D}_s)$$

(by definition of Y_{s+2})

$$= \alpha + \phi E(Y_{s+1}|\tilde{D}_s) + \beta_0 E(X_{s+2}|\tilde{D}_s) + \beta_1 E(X_{s+1}|\tilde{D}_s) + \mathbb{E}(\epsilon_{s+2}|\tilde{D}_s)$$

(by linearity of the conditional expectation)

$$= \alpha + \phi E(Y_{s+1}|\tilde{D}_s) + \beta_0 E(X_{s+2}|\tilde{D}_s) + \beta_1 E(X_{s+1}|\tilde{D}_s)$$

($\mathbb{E}(\epsilon_{s+2}|\tilde{D}_s) = \mathbb{E}(\epsilon_{s+2}) = 0$ because $\{\epsilon_s\}$ is white noise and $\{Y_t\}$ is generated by an ADL with exogenous regressor $\{X_t\}$, hence ϵ_{s+2} is independent of $\tilde{D}_s$)

$$= \alpha + \phi \tilde{Y}_{s+1} + \beta_0 \tilde{X}_{s+2} + \beta_1 \tilde{X}_{s+1}$$

(the final result would be obtained by substituting in the expressions for $\tilde{Y}_{s+1}$, $\tilde{X}_{s+2}$ and $\tilde{X}_{s+1}$ derived above)

(c) (**15pts**) Let $\{Y_t\}_{t\in\mathbb{Z}}$ and $\{X_t\}_{t\in\mathbb{Z}}$ be $I(1)$ time series. Suppose that you have obtained the following estimates for the parameters of the ADL(1,1) model above:

Parameter	α	ϕ	β_0	β_1	σ^2_ϵ
Estimate	0.11	0.94	1.28	0.01	1.14

Furthermore, suppose that the p-values you obtained indicate that all parameters are significantly different from zero at the 5% significance level. Does $\{X_t\}_{t\in\mathbb{Z}}$ Granger cause $\{Y_t\}_{t\in\mathbb{Z}}$? Can $\{Y_t\}_{t\in\mathbb{Z}}$ Granger cause $\{X_t\}_{t\in\mathbb{Z}}$? Are $\{Y_t\}_{t\in\mathbb{Z}}$ and $\{X_t\}_{t\in\mathbb{Z}}$ cointegrated? Justify your answers carefully and in detail.

Answer:

[**2pts**] A time-series $\{X_t\}_{t\in\mathbb{Z}}$ causes a time-series $\{Y_t\}_{t\in\mathbb{Z}}$ if past values of $\{X_t\}_{t\in\mathbb{Z}}$ provide statistically significant information about future values of $\{Y_t\}_{t\in\mathbb{Z}}$.

[**2pts**] If the estimation results and the p-values reported in the table are reliable (for example, if the residuals seem to be white noise), then, at the 5% significance level we could conclude that β_1 is significantly different from zero, and hence $\{X_t\}_{t\in\mathbb{Z}}$ Granger causes $\{Y_t\}_{t\in\mathbb{Z}}$.

[**2pts**] Note however that the answer depends on the adopted significance level. At more stringent significance levels we might conclude that β_1 is not significantly different from zero anymore and thus that $\{X_t\}_{t\in\mathbb{Z}}$ does not Granger cause $\{Y_t\}_{t\in\mathbb{Z}}$.

[**2pts**] Yes, $\{Y_t\}_{t\in\mathbb{Z}}$ can Granger cause $\{X_t\}_{t\in\mathbb{Z}}$. It is possible that past values of $\{Y_t\}_{t\in\mathbb{Z}}$ provide statistically significant information about future values of $\{X_t\}_{t\in\mathbb{Z}}$, regardless of the reverse.

[**3pts**] We can rewrite the ADL in an error correction form.

$$\begin{aligned} & Y_t = \alpha + \phi Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t \\ \Leftrightarrow \quad & \Delta Y_t = \alpha + (\phi - 1) Y_{t-1} + \beta_0 \Delta X_t + (\beta_0 + \beta_1) X_{t-1} \\ \Leftrightarrow \quad & \Delta Y_t = -(1-\phi) \left(Y_{t-1} - \frac{\alpha}{1-\phi} - \frac{\beta_0 + \beta_1}{1-\phi} X_{t-1} \right) + \beta_0 \Delta X_t \end{aligned}$$

[**2pts**] Therefore, if $|\phi| < 1$, then $-2 < -(1-\phi) < 0$. If the estimation results presented in the table are reliable, then the point estimate of 0.94 for ϕ suggests (by Granger's representation theorem) that $\{Y_t\}_{t \in \mathbb{Z}}$ and $\{X_t\}_{t \in \mathbb{Z}}$ are indeed cointegrated.

[**2pts**] Note however that we should test for cointegration using an appropriate test.