FORECASTING & PREDICTIVE ANALYTICS

1.a) the process $\{y_t\}$ is weakly stationary if its first & second moments exist and are time invariant.

- the first moment is $\{E[y_t]\}$
- the second moment consists of $E[\underline{y}_t \underline{y}_t']$ where

$$\underline{y}_t = \begin{bmatrix} \vdots \\ y_t \\ \vdots \end{bmatrix} \text{ infinite vector}$$

i.e the second moment consists of the variances and covariances.

weak stationary is useful as the time series then has a constant distribution [that can be used for forecasting] and also the law of large numbers and central limit theorem can be used so estimation is feasible with standard asymptotics.

b) i. $\Delta y_t = -0.2\, y_{t-1} + \epsilon_t$ AR(1) with coef 0.8 which is less than 1 in modulus $\therefore$ stationary (weakly)

ii. $y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t + \theta \epsilon_{t-1}$ ARMA(2,1) is stationary if roots of AR lag polynomial $A(L) = 1 - \phi_1 L - \phi_2 L^2$ have modulus greater than unity (or one is equal to a root of MA lag polynomial $1+\theta L$ so the ARMA(2,1) is in fact an AR(1).)

iii. $y_t = \phi_0 + 0.1 x_{t-1} + \epsilon_t$; $x_t = x_{t-1} + \epsilon_t$

$x_t \sim I(1)$ non stationary (random walk)

$$y_t = \phi_0 + 0.1 \Big[\underbrace{\frac{y_{t-1} - \phi_0 - \epsilon_{t-1}}{0.1}}_{x_{t-2}} + \epsilon_{t-1} \Big] + \epsilon_t$$

$\therefore \quad y_t = \phi_0 (1-1) + y_{t-1} + \epsilon_t + 0.1 (1 - \frac{1}{0.1}) \epsilon_{t-1}$

$= y_{t-1} + \epsilon_t - 0.9 \epsilon_{t-1}$

$\sim$ ARIMA (1, 1) with an autoregressive unit root

$\therefore$ non stationary.

iv. AR(1) with coefficient 0.8 : same process as i. $\therefore$ stationary

v. $y_t = y_{t-1} + \epsilon_t - \epsilon_{t-1}$: $(1-L) y_t = (1-L) \epsilon_t$ $\therefore$ $y_t = \epsilon_t \sim$ iid

$\therefore$ stationary (white noise)

vi. $y_t = y_{t-2} + \epsilon_t$ $\therefore$ $(1 - L^2) y_t = \epsilon_t$

autoregressive unit root (1 and -1)

$\therefore$ non stationary.

---

2. AR(2) $\quad y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$

a) $\Delta y_t = (\phi_1 - 1) y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$

$= (\phi_1 - 1) y_{t-1} - \phi_2 \Delta y_{t-1} + \phi_2 y_{t-1} + \epsilon_t$

$= (\phi_1 + \phi_2 - 1) y_{t-1} - \phi_2 \Delta y_{t-1} + \epsilon_t \qquad (*)$

b) for $(*)$ to become an AR(1) in Δy_t we require

$\phi_1 + \phi_2 - 1 = 0$ then $-\phi_2 = \phi_1 - 1$

$\Delta y_t = (\phi_1 - 1) \Delta y_{t-1} + \epsilon_t$

c) when $\phi_1 + \phi_2 - 1 = 0$ then Δy_t can, or not, be stationary

but y_t is not - if $|\phi_1 - 1| < 1$ then $\Delta y_t \sim I(0)$; $y_t \sim I(1)$

d) see the lecture notes : the ADF is a test that $\phi_1 + \phi_2 - 1 = 0$

e) $y_{t+1|t} = \phi_1 y_t + \phi_2 y_{t-1}$ $\therefore$ $e_{t+1|t} = \epsilon_{t+1}$

$y_{t+2|t} = \phi_1 y_{t+1|t} + \phi_2 y_t = \phi_1^2 y_t + \phi_1 \phi_2 y_{t-1} + \phi_2 y_t$

$= (\phi_1^2 + \phi_2) y_t + \phi_1 \phi_2 y_{t-1}$ with $y_{t+2} = y_{t+2|t} + \epsilon_{t+2} + \phi_1 \epsilon_{t+1}$

$\therefore$ $e_{t+2|t} = \epsilon_{t+2} + \phi_1 \epsilon_{t+1}$

- $e_{t+1|t} = \epsilon_{t+1}$ so the process $\{e_{t+1|t}\}_t$ is iid
  the ACF is zero, so is the PACF
- $e_{t+2|t} = \epsilon_{t+2} + \phi_1 \epsilon_{t+1}$ so the process is an MA(1)
  i.e the ACF ρ_k is non zero for ρ_1 and $\rho_k = 0 ; k > 1$
  the PACF decays exponentially towards zero.

MULTIVARIATE MODELS

3. VAR(1) $$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} .4 & .3 \\ .8 & .6 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix}$$

$$\begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix} \overset{iid}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon\eta} \\ \sigma_{\epsilon\eta} & \sigma_\eta^2 \end{bmatrix} \right)$$

a) we look at the eigenvalues ~~of the char. polynom~~ of the matrix $\begin{bmatrix} .4 & .3 \\ .8 & .6 \end{bmatrix}$

i.e. s.t $$\left| \begin{bmatrix} .4 & .3 \\ .8 & .6 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

$$\therefore \quad \begin{vmatrix} .4 - \lambda & .3 \\ .8 & .6 - \lambda \end{vmatrix} = 0$$

$$(0.4 - \lambda)(0.6 - \lambda) - 0.8 \times 0.3 = 0$$

$$\lambda^2 - \lambda + 0.24 - 0.24 = 0$$

$$\lambda(\lambda - 1) = 0$$

the two eigenvalues are 0 and 1

1 being an eigenvalue, the process is not stationary.

b)

b) $$x_t - \beta y_t = \begin{bmatrix} 1 \\ -\beta \end{bmatrix}' \begin{bmatrix} x_t \\ y_t \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ -\beta \end{bmatrix}' \left(\begin{bmatrix} .4 & .3 \\ .8 & .6 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix} \right)$$

$$= \begin{bmatrix} .4 - .8\beta & .3 - .6\beta \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \epsilon_t - \beta \eta_t$$

$$= (1-2\beta) \left[.4 x_{t-1} - .3 y_{t-1} \right] + \epsilon_t - \beta \eta_t$$

$$= 0.4 (1-2\beta) \left(x_{t-1} - \frac{3}{4} y_{t-1} \right) + \epsilon_t - \beta \eta_t$$

* so the first solution is to assume $x_{t-1} - \frac{3}{4} y_{t-1} = x_{t-1} - \beta y_{t-1}$

i.e $\beta = 3/4$

then $$\begin{cases} x_t - \frac{3}{4} y_t = -0.2 \left(x_{t-1} - \frac{3}{4} y_{t-1} \right) + u_t \\ u_t = \epsilon_t - \frac{3}{4} \eta_t \sim iid \end{cases}$$

* the second solution is to assume $1 - 2\beta = 0$, i.e $\beta = \frac{1}{2}$

then $x_t - \frac{1}{2} y_t = \epsilon_t - \beta \eta_t \sim iid$

this is a "trivial" AR(1) in the sense that the autoregressive coefficient is zero.

c) cointegration arises when some processes are non stationary and integrated of order d, but there exists a linear combination of these processes that is integrated of order d-1.

here given that $\begin{bmatrix} x_t \\ y_t \end{bmatrix}$ is non stationary but there exists a linear combination that is stationary

($x_t - \frac{3}{4} y_t$ is stationary, and so is $x_t - 1/2\, y_t$)

then x_t and y_t cointegrate.

d) to compute $\frac{\partial x_{t+k}}{\partial \epsilon_t}$ we need to find a way to compute changes to ϵ_t that do not affect η_t or else it is not truly a partial derivative.

if $\sigma_{\epsilon\eta} = 0$ so ϵ_t and η_t do not correlate then $\partial x_{t+k} / \partial \epsilon_t$ is easy to compute or else we also need to take into account $\frac{\partial \eta_t}{\partial \epsilon_t}$

EMPIRICAL ANALYSIS

4. a) Figures 1 and 2 present Sales and Advert in levels and in logs so the stationarity properties are the same for the two figures - we focus on fig. 2.

- the processes do not look stationary from the figure as they do not cross often their sample average (although the latter seems "stable" over the sample). The ACF do exhibit some persistence (linear decay) but they might just reflect strong stationary persistence.

- by contrast figure 3 shows processes that seem stationary - the ACF confirm this (possibility of changing volatility though)

b) based on the figures, we choose $d=1$ for both variables

- for LAdvert we look at the ACF / PACF of DLAdvert - ~~there's~~ there isn't much autocorrelation, except maybe for ACF/PACF at lag 2
  → looks like an MA(2), i.e $p=0, q=2$
  we may want to start with $q=4$ for safety.

- for LSales (i.e looking at DLSales) the possibility of an AR(1) cannot be excluded (decay in ACF) so we may want to start with ARMA(1,1), i.e $p=q=1$

c) for DLAdvert, an AR(2) was estimated - we see the coefficient on $DLAdvert_{t-2}$ is significant but not the one on $DLAdvert_{t-1}$ → the latter could be removed (also the intercept is not significant)

we had recommended an MA(2) but it seems an AR(2) also fits the data (we should check further)

d) the role of the dummy variable is to ensure a perfect fit in 1936 so that the outlier does not impair the estimation of other parameters.
we see that it impacts in particular the coefficient on $DLAdvert_{t-1}$ which becomes almost significant.

e) I suggest using the model in d)

a question is whether we should use $DLAdvert_{t-1}$ or not in the forecasts. Given that we cannot reestimate the model without it, I suggest to keep it.

then the forecasts are

$$\widehat{DLAdvert}_{1937|1936} = 0.02 - 0.22 \cdot DLAdvert_{1936} - 0.28\, DLAdvert_{1935}$$

$$= 0.385$$

so $\widehat{LAdvert}_{1937|36} = LAdvert_{1936} + \widehat{DLAdvert}_{1937|36}$

$$= 5.82 + 0.385 = 6.21$$

and for $\widehat{DLAdvert}_{1938|36} = 0.02 - 0.22\, \widehat{DLAdvert}_{1937|36} - 0.28\, DLAdvert_{1935}$

$$= 0.02 - 0.22 \times 0.385 + 0.28 \times 0.87$$

$$= 0.18$$

so $\widehat{LAdvert}_{1938|36} = \widehat{LAdvert}_{1937|36} + \widehat{DLAdvert}_{1938|36}$

$$= 6.21 + 0.18$$

$$= 6.39.$$

this is the iterated multi-step ahead forecast.

f) we obtain $\widetilde{LAdvert}_{1937|36} = 5.45$

$\widetilde{LAdvert}_{1938|36} = 6.71$

~~these are direct fcsts for 1937~~

g) iterated is more efficient if the model is well specified, direct is more robust if model is misspecified.

here the 2-step direct model would be obtained ~~by~~ by the model of d) where we set the coef on $DLAdvat_{t-1}$ to zero. (since it is not significant)

→ we see that modeling D2LAdvert as in f) seem to generate a better fit when we compare Figures 5 and 6.

h) Granger non causality arises when a variable does not enter conditional distributions

e.g. y_t does not Granger cause x_t if

$$D(x_t \mid x_{t-1}, x_{t-2}, \dots ; y_{t-1}, y_{t-2}, \dots)$$

$$= D(x_t \mid x_{t-1}, x_{t-2} \dots)$$

we can specialize it to conditional expectations, variances...

here we see that DLAdvert does not G.C. DLSales since the coef on $DLAdvat_{t-1}$ is not significant in the second equation.

by contrast DLSales does G.C. DLAdvert.