Question 2. French wine. The researcher is interested in studying how the relative prices of wine influence its consumption in France. She's considering two alternative regression models - simple linear regression (1) and model in differences (2) - using data on wine consumption W_t (in billions of euro) subject to the relative price index of wine, P_t on the base of annual data (1995-2018).

$$W_t = \alpha_1 + \alpha_2 P_t + u_t \qquad (1)$$

$$\Delta W_t = \gamma_1 + \gamma_2 \Delta P_t + v_t \qquad (2\ )$$

where $W_t = W_t - W_{t-1}$, $\Delta P_t = P_t - P_{t-1}$, and u_t, and v_t are disturbance terms.
Estimating these equations she obtained

$$\hat{W}_t = 15.6 - 0.03 P_t \qquad R^2 = 0.03$$
$$(3.86)\ (0.04) \quad DW = 0.11 \quad RSS = 22.4 \qquad (1e)$$

$$\Delta\hat{W}_t = 0.13 - 0.06 \Delta P_t \qquad R^2 = 0.29$$
$$(0.06)\ (0.02) \quad DW = 3.02 \quad RSS = 1.95 \qquad (2e)$$

(a) (10 marks) The colleague advised the researcher to check the data series for stationarity by telling her about the computer experiments of Newbold-Granger, and said that probably her estimated regression (1e) is spurious regression.

□ What is stationarity? Why it is important?

Solution: A process $\{y_t\}$ is (covariance) stationary if its mean and variance exist and do not depend on time and its covariance is a function of distance in time only (not location). Violation of either of these requirements renders the process non-stationary. This can lead to the fact that the estimates of the coefficients and standard errors become inconsistent, the tests are invalid, often this leads to the appearance of an spurious regression

□ Explain why spurious regression can occur if the condition of stationary of time series is violated?

Solution: Regression is called **spurious regression** if in fact there is no relationship between two variables but it has some signs of high-quality regression: high R-square value, significant coefficients. Basically, such a result can never be excluded (first type error), but it is characteristic of the spurious regression that at repetition of calculations on various samples such a situation occurs abnormally often**..**
The possible mechanism of emerging spurious regression is as follows. If we regress Y_t on X_t, i.e.

$$Y_t = \pi_0 + \pi_1 X_t + v_t,$$

Granger and Newbold have shown that although there is no relationship between *Y* and *X*, the regression will produce a *t*-ratio which **will reject the null hypothesis** $H_0 : \pi_1 = 0$.
The reason for this result is that if $H_0 : \pi_1 = 0$, then

$$Y_t = \pi_0 + v_t$$

and since Y_t is $I(1)$ and π_0 is constant, it follows that v_t must be $I(1)$. This violates the standard distributional theory based on the assumption that v_t is stationary, i.e. v_t is $I(0)$. Hence the misleading result.

(b) (10 marks) The researcher decided to test the data series for stationarity by constructing the following Dickey-Fuller equations

$$\hat{\Delta W}_t = 5.48 - 0.50W_{t-1} + 0.07t \quad R^2 = 0.25 \qquad (3e)$$
$$(2.17)\ (0.20)\quad (0.03)$$

$$\hat{\Delta P}_t = 5.48 - 0.05P_{t-1} + 0.46\Delta P_{t-1} - 0.08t \quad R^2 = 0.43 \qquad (5e)$$
$$(8.09)\ (0.09)\quad (0.25)\quad (0.08)$$

$$\hat{\Delta^2 W}_t = -1.15\Delta W_{t-1} \quad R^2 = 0.58 \qquad (4e)$$
$$(0.23)$$

$$\hat{\Delta^2 P}_t = -2.16\Delta P_{t-1} \quad R^2 = 0.19 \qquad (6e)$$
$$(0.18)$$

□ Interpret the results of the evaluation of Dickey-Fuller equations. Which of the time series are stationary and which are non-stationary?

Solution: Using t-ADF table.

For equation (3e) $t_{W_{t-1}} = \dfrac{b}{s.e.(b)} = \dfrac{-0.50}{0.2} = -2.5 > -3.62 = t^{5\%}_{ADF,\,crit}(23, \text{trend in model})$ - wine consumption y_t is non-stationary, there is also significant trend.

For equation (4e) $t_{\Delta W_{t-1}} = \dfrac{b}{s.e.(b)} = \dfrac{-1.15}{0.23} = -5 < -3.77 = t^{1\%}_{ADF,\,crit}(22, \text{no trend})$ - their differences Δy_t are stationary.

Equation (5e) uses additional lag and shows that prices P_t is obviously non-stationary but has no trend.
t-statistic in equation (6e) indicates that ΔP_t is stationary.

□ What is the difference between the specification of equation (3e) and the specification of equation (5e), does each of these specifications have any advantages (and possible disadvantages)?

Solution: The test is sensitive to the presence of serial correlation in the error term so we need to take steps to remove the effects of this serial correlation - this is done by including lagged values of W_t in the regression. On the other hand equation (5e) has reduced number of degrees of freedom, so the estimators become less efficient.

(c) (10 marks) □ Explore Dickey-Fuller equation models mathematically. Derive theoretical model corresponding to the time series model (3e). State null and alternative hypothesis and decision rule. Explain the logic of Dickey-Fuller test.

Solution: The standard test for a unit root is due to Dickey and Fuller and is based on the model $W_t = \beta_1 + \beta_2 W_{t-1} + \gamma t + u_t$ which can be re-written as $\Delta W_t = \beta_1 + (1-\beta_2)W_{t-1} + \gamma t + u_t$ where $\Delta y_t = y_t - y_{t-1}$. The null hypothesis for stationarity is $H_0 : \beta_2 - 1 = 0$, $H_A : \beta_2 - 1 < 0$. We cannot use the standard t-test procedure in this case because the distribution of the t-statistic is not a t-distribution so critical values have been computed by Dickey and Fuller using Monte Carlo techniques.

Solution: we start now from the model $P_t = \beta_1 + \beta_2 P_{t-1} + \beta_3 P_{t-2} + \gamma t + u_t$ to control for AR(1) serial correlation. This is more easily tested by using the model $\Delta P_t = \beta_1 + (1-\beta_2-\beta_3)P_{t-1} - \beta_3 \Delta P_{t-2} + \gamma t + u_t$ with null hypothesis $H_0 : 1 - \beta_1 - \beta_2 = 0$. Once again use Dickey-Fuller tables.