THE UNIVERSITY OF MELBOURNE

Student ID ________________

Semester Two Assessment, 2023

Faculty / Dept: **Faculty of Business and Economics / Dept. of Economics**

Subject Number **ECON90033**

Subject Name **Quantitative Analysis of Finance I**

Writing time 2 hrs

Reading 15 minutes

Open Book status Yes

Number of pages (including this page) 16 pages, 1 front page, 15 pages of questions

Authorised Materials: Calculators: Casio FX82 (any suffix).
Any printed or handwritten material.

Instructions to Students:

This examination paper contributes **60 per cent** of the assessment in ECOM90033.

This exam has FOUR (4) questions, each of them consisting of several parts and tasks. Every question, part and task is compulsory. Answer ALL questions in the script books provided.

The marks allocated for each question are:

Question 1: marks; Question 2: marks; Question 3: marks; Question 4: marks

Total: 60 marks

Suggested time allocation: Question 1, minutes
Question 2, minutes
Question 3, minutes
Question 4, minutes

Instructions to Invigilators: Students are to be supplied with examination SCRIPT BOOKS.
This exam paper may not be taken from the examination room.
This is an open book exam.

Paper to be held by Baillieu Library: no

Extra Materials required: none.

The exam will be an open-book exam, you will be allowed to use your textbook(s), study notes and other printed or handwritten materials while you take the exam. A formula sheet and statistical tables will not be provided on the exam, it is your responsibility to bring them, and your Casio FX82 calculator as well.

There will be four questions in the final exam paper, each of them consisting of several parts tasks. Every question, part and task is compulsory. This sample paper also has four questions, and it is similar to the final exam paper in terms of style, length and difficulty. Note, however, that the four questions in this sample exam paper are not meant to cover all examinable topics and the four questions in the final exam paper might be related to different topics.

Event analysis is widely used in empirical finance to model the effect of some discrete event, like the announcement of the change in a company's chief executive officer (CEO), on financial variables. In order to do so, the overall event is decomposed into three sub-events: the part that is anticipated by the market, the part that occurs at the time of the event, and the part that happens after the event has occurred. This is achieved by specifying a regression equation that represents 'normal' market returns, and then modifying this equation to account for these three sub-events through the inclusion of indicator (dummy) variables.

As an example for event analysis, consider Exxon (ExxonMobil since 1999), an American multinational oil and gas corporation. Lee Roy Raymond was its CEO from 1993 to 2005 and on his retirement in December 2005 he received the largest retirement package ever recorded of around $400 million. How did the markets view this event? Questions 1 and 2 in this exam paper serve to answer this question.

Question 1 (marks =)

Exxon event study, part 1.

Monthly observations are collected on the equity price of Exxon and ExxonMobil (*EX*) and the S&P 500 index (*SP*) for the period January 1970 to March 2010.

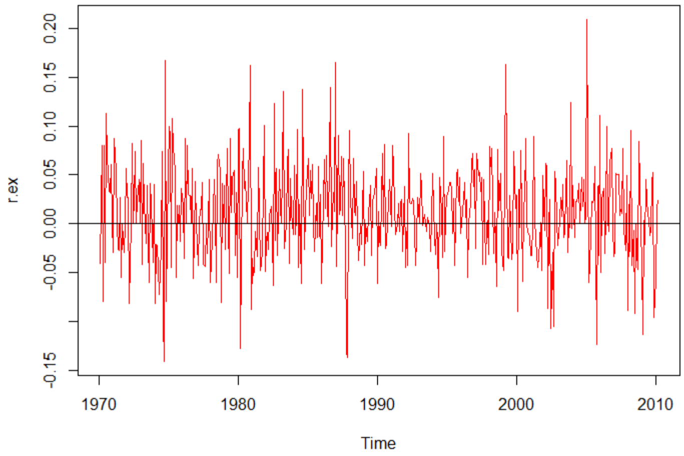
a) From EX_t and SP_t two new series are derived by executing the following *R* commands:

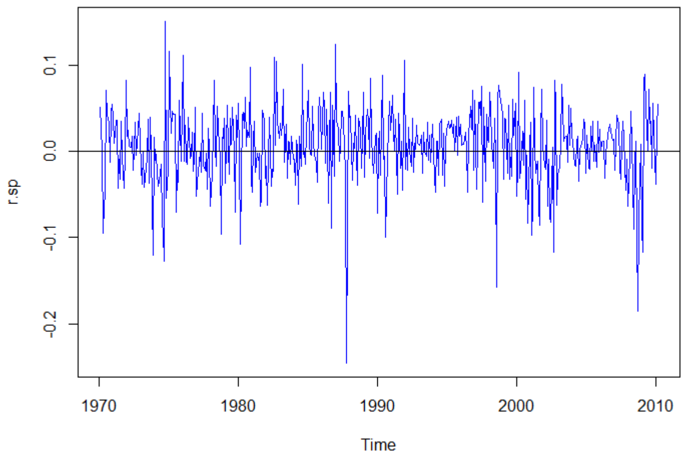
r.ex = diff(log(EX))
r.sp = diff(log(SP))

What are these new variables and what do they measure?

r.ex and r.sp are the log returns on Exxon and on the S&P 500 index. They well approximate the simple returns, as long as the simple returns are relatively small.

b) The figures on the next page illustrate the two new series, $r.ex_t$ and $r.sp_t$. Both series appear stationary and unit root / stationarity tests on their levels and first differences confirm that they are indeed I(0) variables.





What does the statement, "$r.ex_t$ and $r.sp_t$ are I(0) variables", actually mean?

Suppose for a moment, that $r.ex_t$ and $r.sp_t$ are I(1) variables. What would this imply on a regression of $r.ex_t$ and $r.sp_t$?

"$r.ex_t$ and $r.sp_t$ are I(0) variables" means that $r.ex_t$ and $r.sp_t$ are integrated of order zero, implying that they do not need to be differenced at all to achieve stationarity as they are stationary variables.

If $r.ex_t$ and $r.sp_t$ are I(1) variables, a regression of $r.ex_t$ and $r.sp_t$ would be spurious, i.e., the OLS sample regression equation might look good (high R^2, significant F- and t-statistics), but it would not have any real meaning because there is no direct relationship between the variables.

c) A simple linear regression of $r.ex_t$ and $r.sp_t$ is estimated and subjected to three diagnostic tests. The relevant *R* commands and printouts are on the next page.

You have three tasks. Explain your answers and report every numerical value from the printouts that your answers rely on.

i. Comment on the adjusted R^2 statistic and on the *F*-test for the overall significance.

ii. Interpret the point estimates of the intercept and slope parameters in the context of the estimated model.

iii. What are the purposes of the three diagnostic tests and what conclusions do you draw from them? Be precise.

```
> m1 = lm(r.ex ~ r.sp)
> summary(m1)

Call:
lm(formula = r.ex ~ r.sp)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.121015 -0.026968 -0.001665  0.024257  0.189657 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.008213   0.001916   4.286 2.19e-05 ***
r.sp        0.620721   0.042090  14.747  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04176 on 480 degrees of freedom
Multiple R-squared:  0.3118,	Adjusted R-squared:  0.3104 
F-statistic: 217.5 on 1 and 480 DF,  p-value: < 2.2e-16


> library(lmtest)
> bgtest(m1, order = 12, type = "Chisq")

	Breusch-Godfrey test for serial correlation of order up to 12

data:  m1
LM test = 16.28, df = 12, p-value = 0.1787


> library(FinTS)
> ArchTest(m1.res, lags = 12)

	ARCH LM-test; Null hypothesis: no ARCH effects

data:  m1.res
Chi-squared = 14.115, df = 12, p-value = 0.2934
```

```
> resettest(m1, power = 3, type = "fitted")

        RESET test

data:  m1
RESET = 0.10382, df1 = 1, df2 = 479, p-value = 0.7474
```

i. *The adjusted R^2 statistic is 0.3104. It means that this simple linear regression can account for about 31% of the total sample variation in the log returns to Exxon.*

   The F-statistic is 217.5 with a p-value of 2.2e-16 = 2.2. / 10^{16}, which is practically zero. Hence, the null hypothesis that the slope parameter is zero is rejected at any reasonable significance level in favour of the alternative hypothesis that the slope parameter is different from zero.

ii. *The point estimate of the intercept is 0.008213. It means that when the log return to the S&P 500 index is zero, the expected log return to Exxon is about 0.0082, i.e., 0.82%.*

   The point estimate of the slope parameter is about 0.621. It means that a 1 percentage point increase of the return on the S&P 500 index is expected to be accompanied by an 0.621 percentage point increase of the return to Exxon.

iii. *The first test is the BG test for serial correlation of order up to 12. The objective is to find out whether the error term satisfied the null hypothesis of no autocorrelation of orders 1 to 12. The p-value is 0.1787 > 0.10, so the null hypothesis is maintained even at the 10% significance level. Hence, there is no evidence of autocorrelation of orders 1 to 12.*

   The second test is the ARCH LM test. The null hypothesis is that there are no ARCH effects of orders 1 to 12, while the alternative hypothesis is that there is (are) some ARCH effect(s). The p-value is 0.2934 > 0.10, so the null hypothesis is maintained even at the 10% significance level. Hence, there is no evidence of ARCH effects of orders 1 to 12.

   The third test is the RESET test. The null hypothesis is that the estimated regression model has the correct functional form, while according to the alternative hypothesis the functional form is incorrect. The p-value is 0.7474 > 0.10, so the null hypothesis is maintained even at the 10% significance level. Hence, there is not sufficient evidence to conclude that the functional form is incorrect.

Question 2 (marks =)

Exxon event study, part 2.

a) To determine how the market viewed Lee Roy Raymond's retirement, the following multiple linear regression model is specified:

$$r.ex_t = \beta_0 + \beta_1 r.sp_t + \delta_1 I_{1t} + \delta_2 I_{2t} + \delta_3 I_{3t} + \delta_4 I_{4t} + \delta_5 I_{5t} + \varepsilon_t$$

where I_{1t}, I_{2t}, I_{3t}, I_{4t}, I_{5t} are indicator (dummy) variables defined around the retirement of Lee Roy Raymond in December 2005 as

$$I_{1t} = \begin{cases} 1: & \text{Oct-2005} \\ 0: & \text{otherwise} \end{cases} , \quad I_{2t} = \begin{cases} 1: & \text{Nov-2005} \\ 0: & \text{otherwise} \end{cases} ,$$

$$I_{3t} = \begin{cases} 1: & \text{Dec-2005} \\ 0: & \text{otherwise} \end{cases} ,$$

$$I_{4t} = \begin{cases} 1: & \text{Jan-2006} \\ 0: & \text{otherwise} \end{cases} , \quad I_{5t} = \begin{cases} 1: & \text{Feb-2006} \\ 0: & \text{otherwise} \end{cases}$$

Hence, I_{1t} and I_{2t} refer to the two months right before the retirement (pre-event), I_{3t} to the month of the retirement (actual event), and I_{4t} and I_{5t} refer to the two months straight after the retirement (post-event). The δ_1, δ_2, δ_3, δ_4, δ_5 parameters of these indicator variables measure the expected abnormal return associated with the event in the 5 months of the event window.

This augmented multiple linear regression model is estimated from the same data as the simple linear regression model in Question 1. The relevant *R* commands and printout are on the next page.

You have two tasks. Explain your answers and report every numerical value from the printouts that your answers rely on.

i. Compare the adjusted R^2 statistic of this model to the adjusted R^2 statistic of the model in Question 1.

ii. Interpret the point estimates that are significant at the 5% level.

```
Call:
lm(formula = r.ex ~ r.sp + I1 + I2 + I3 + I4 + I5)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.115389 -0.026772 -0.001619  0.023594  0.189469 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.008502   0.001904   4.466 9.97e-06 ***
r.sp         0.615293   0.041652  14.772  < 2e-16 ***
I1          -0.121402   0.041343  -2.936  0.00348 ** 
I2           0.008451   0.041350   0.204  0.83815    
I3          -0.040512   0.041333  -0.980  0.32751    
I4           0.086872   0.041340   2.101  0.03613 *  
I5          -0.058925   0.041333  -1.426  0.15463    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04129 on 475 degrees of freedom
Multiple R-squared:  0.3343,	Adjusted R-squared:  0.3259 
F-statistic: 39.76 on 6 and 475 DF,  p-value: < 2.2e-16
```

i. *The adjusted R^2 statistic of this multiple linear regression model is 0.3259, only slightly larger than the adjusted R^2 statistic of the simple linear regression model in Question 1. This means that the contribution of the five indicator variables is minimal and thus the market's reaction to Lee Roy Raymond's retirement might be negligible.*

ii. *At the 5% significance level only the intercept and the slopes of r.sp, I_1 and I_4 are significant.*

The point estimate of the intercept is 0.008502. It means that when the log return to the S&P 500 index is zero and each dummy variable is zero, i.e., outside the 5-month event window, the expected log return to Exxon is about 0.0085, i.e., 0.85%.

The slope estimate of I_1 is -0.121402. It means that for any given r.sp value, the log return to Exxon is expected to be about 0.1214 lower two months before Lee Roy Raymond's retirement than in any other month of the sample period.

The slope estimate of I_4 is 0.086872. It means that for any given r.sp value, the log return to Exxon is expected to be about 0.0869 higher one month after Lee Roy Raymond's retirement than in any other month of the sample period.

b) Two tests are performed to evaluate the overall significance and the net effect of Lee Roy Raymond's retirement on the market. The *R* commands and printouts for the two tests are below and on the next page.

Test 1:

```
> library(car)
> linearHypothesis(model = m2, c("I1", "I2", "I3", "I4", "I5"))
Linear hypothesis test

Hypothesis:
I1 = 0
I2 = 0
I3 = 0
I4 = 0
I5 = 0

Model 1: restricted model
Model 2: r.ex ~ r.sp + I1 + I2 + I3 + I4 + I5

  Res.Df     RSS Df Sum of Sq      F   Pr(>F)
1    480 0.83717
2    475 0.80976  5  0.027408 3.2155 0.007252 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test 2:

```
> linearHypothesis(model = m2, c("I1 + I2 + I3 + I4 + I5 = 0"))
Linear hypothesis test

Hypothesis:
I1  + I2  + I3  + I4  + I5 = 0

Model 1: restricted model
Model 2: r.ex ~ r.sp + I1 + I2 + I3 + I4 + I5

  Res.Df     RSS Df Sum of Sq     F Pr(>F)
1    476 0.81288
2    475 0.80976  1  0.003118 1.829 0.1769
```

You have three tasks. Explain your answers and report every numerical value from the printouts that your answers rely on.

i. Briefly evaluate the printout of Test 1 at the 5% significance level. Clearly state the hypotheses, the statistical decision, and the conclusion in the context of this case study.

ii. Briefly evaluate the printout of Test 2 at the 5% significance level. Clearly state the hypotheses, the statistical decision, and the conclusion in the context of this case study.

iii. Did the retirement of Lee Roy Raymond have a significant net effect on the market in the 5-month event window?

i. *In test 1 the null hypothesis is that each dummy variable has zero slope parameter, and the alternative hypothesis is that some of them has (have) nonzero slope parameter(s). The p-value is 0.007252 < 0.05, so the null hypothesis is rejected at the 5% significance level, implying that some dummy variable(s) has (have) nonzero slope parameter and thus Lee Roy Raymond's retirement had some significant impact on the log return to Exxon in the event window.*

ii. *In test 2 the null hypothesis is that the slope parameters of the five dummy variables add up to zero, while the alternative hypothesis is that their sum is different from zero. The p-value is 0.1769 > 0.05, so the null hypothesis is maintained at the 5% significance level, implying that the sum of the slope estimate of the five dummy variables is only insignificantly different from zero.*

iii. *The net effect of Lee Roy Raymond' retirement in the 5-month event window can be captured by the sum of the slope estimates of the five dummy variables. In part (ii) above it was already concluded that this sum is not significantly different from zero. Hence, Lee Roy Raymond's retirement had an insignificant net effect on the log return to Exxon in the event window.*

c) Consider the 5-month event window from October 2005 to February 2006, centered around the retirement of Lee Roy Raymond in December 2005. The observed values of $r.sp_t$ over this time are

	Oct-05	Nov-05	Dec-05	Jan-06	Feb-06
r.sp	-0.0179	0.0346	-0.0010	0.0251	-0.0005

Substitute these values into the sample regression equation in part (a) and calculate the point estimates of the log return to Exxon over this period.

The sample regression equation is

$$\widehat{r.ex_t} = \hat{\beta}_0 + \hat{\beta}_1 r.sp_t + \hat{\delta}_1 I_{1t} + \hat{\delta}_2 I_{2t} + \hat{\delta}_3 I_{3t} + \hat{\delta}_4 I_{4t} + \hat{\delta}_5 I_{5t}$$
$$= 0.0085 + 0.6153 r.sp_t - 0.1214 I_{1t} + 0.0085 I_{2t} - 0.0405 I_{3t} + 0.0869 I_{4t} - 0.0589 I_{5t}$$

The point estimates of the log return to Exxon are the following.

October 2005: $\widehat{r.ex_t} = 0.0085 + 0.6153 \times (-0.0179) - 0.1214 = -0.1239$

November 2005: $\widehat{r.ex_t} = 0.0085 + 0.6153 \times 0.0346 + 0.0085 = 0.0383$

December 2005: $\widehat{r.ex_t} = 0.0085 + 0.6153 \times (-0.0010) - 0.0405 = 0.0330$

January 2006: $\widehat{r.ex_t} = 0.0085 + 0.6153 \times 0.0251 + 0.0869 = 0.1108$

February 2006: $\widehat{r.ex_t} = 0.0085 + 0.6153 \times (-0.0005) - 0.0589 = -0.0507$

Question 3 (marks =)

According to Fomby and Hirschberg (1989), there are two popular economic wisdoms about the economy of Texas USA, namely that "As the oil patch goes, so goes the Texas economy" and "Where the national economy goes, Texas economy need not follow". To verify these beliefs, the authors developed a three-variable VAR model of the Texas economy for the period 1974 Q1 to 1988 Q1.

As a follow up study, an economics honours student collects quarterly data from 1986 Q1 to 2023 Q2 on the following variables:

COP: Crude oil price (West Texas Intermediate, dollars per barrel),
TXNFE: Total number of nonfarm employees in Texas (thousands of persons),
USNFE: Total number of nonfarm employees in the US (thousands of persons).
GNPDEF: Implicit price deflator of US gross national product (2017 = 100).

a) From these data the student generates five new variables by executing the following R commands:

RUSNFE = USNFE - TXNFE
*RCOP = COP/GNPDEF * 100*
*RCP = (RCOP / lag(RCOP,-1) - 1) * 100*
*TXP = (TXNFE / lag(TXNFE,-1) - 1) * 100*
*RUSP = (RUSNFE / lag(RUSNFE,-1) - 1) * 100*

Describe with a few words what these new variables are, what they actually measure.

Th first new variable is the total number of nonfarm employment in the US except the state of Texas. The second variable is the real price crude oil (nominal price divided by the implicit price deflator). The third, fourth and fifth variables are the percentage quarterly changes of the real price of crude oil, the total number of nonfarm employment in the state of Texas, and the total number of nonfarm employment in all other states of the US, respectively.

b) Next, the student performs *ADF* and *KPSS* tests with *R*. The *R* commands and parts of the corresponding printouts are shown on the next four pages.

You have two tasks.

i. The *ADF* and *KPSS* tests in general are popular in practice because they are said to complement each other. In what sense do they complement each other? Briefly explain your answer.

These two tests nicely complement each other because the ADF test is for H_0: unit root vs. H_A: stationarity, while the KPSS test is for H_0: stationarity vs. H_A: unit root.

ii. Evaluate the test results at the 10% significance level. What are the orders of integration of the three variables (*RCP*, *TXP*, *USP*)? Report every numerical value from the printout that your answers rely on.

```
> adf.RCP = ur.df(RCP, type = "drift", selectlags = "BIC", lags = 5)
> summary(adf.RCP)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression drift


Value of test-statistic is: -9.0408

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.46 -2.88 -2.57
```

The test statistic (-9.0408) is smaller than the 10% critical value (-2.57), so H_0 is rejected at the 10% level, implying that RCP is stationary.

```
> adf.DRCP = ur.df(diff(RCP), type = "none", selectlags = "BIC", lags = 5)
> summary(adf.DRCP)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Value of test-statistic is: -9.4839

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

The test statistic (-9.4839) is smaller than the 10% critical value (-1.62), so H_0 is rejected at the 10% level, so the first difference of RCP is stationary.

```
> adf.TXP = ur.df(TXP, type = "drift", selectlags = "BIC", lags = 5)
> summary(adf.TXP)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression drift


Value of test-statistic is: -4.4689

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.46 -2.88 -2.57
```

The test statistic (-4.4689) is smaller than the 10% critical value (-2.57), so H_0 is rejected at the 10% level, implying that TXP is stationary.

```
> adf.DTXP = ur.df(diff(TXP), type = "none", selectlags = "BIC",
+                  lags = 5)
> summary(adf.DTXP)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Value of test-statistic is: -16.1282

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

The test statistic (-16.1282) is smaller than the 10% critical value (-1.62), so H_0 is rejected at the 10% level, so the first difference of TXP is stationary.

```
> adf.RUSP = ur.df(RUSP, type = "drift", selectlags = "BIC", lags = 5)
> summary(adf.RUSP)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression drift

Value of test-statistic is: -4.8155

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.46 -2.88 -2.57
```

The test statistic (-4.8155) is smaller than the 10% critical value (-2.57), so H_0 is rejected at the 10% level, implying that RUSP is stationary.

```
> adf.DRUSP = ur.df(diff(RUSP), type = "none", selectlags = "BIC",
+                   lags = 5)
> summary(adf.DRUSP)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Value of test-statistic is: -18.3435

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

The test statistic (-18.3435) is smaller than the 10% critical value (-1.62), so H_0 is rejected at the 10% level, so the first difference of RUSP is stationary.

```
> kpss.RCP = ur.kpss(RCP, type = "mu")
> summary(kpss.RCP)

#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 4 lags.

Value of test-statistic is: 0.0454

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

The test statistic (0.0454) is smaller than the 10% critical value (0.347), so H_0 is maintained at the 10% level, implying that RCP might be stationary.

```
> kpss.DRCP = ur.kpss(diff(RCP), type = "mu")
> summary(kpss.DRCP)

#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 4 lags.

Value of test-statistic is: 0.0257

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

The test statistic (0.0257) is smaller than the 10% critical value (0.347), so H_0 is maintained at the 10% level, implying that the first difference of RCP might be stationary.

```
> kpss.TXP = ur.kpss(TXP, type = "mu")
> summary(kpss.TXP)

#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 4 lags.

Value of test-statistic is: 0.0531

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

The test statistic (0.0531) is smaller than the 10% critical value (0.347), so H_0 is maintained at the 10% level, implying that TXP might be stationary.

```
> kpss.DTXP = ur.kpss(diff(TXP), type = "mu")
> summary(kpss.DTXP)

#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 4 lags.

Value of test-statistic is: 0.0152

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

The test statistic (0.0152) is smaller than the 10% critical value (0.347), so H_0 is maintained at the 10% level, implying that the first difference of TXP might be stationary.

```
> kpss.RUSP = ur.kpss(RUSP, type = "mu")
> summary(kpss.RUSP)

#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 4 lags.

Value of test-statistic is: 0.1521

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

The test statistic (0.1521) is smaller than the 10% critical value (0.347), so H_0 is maintained at the 10% level, implying that RUSP might be stationary.

```
> kpss.DRUSP = ur.kpss(diff(RUSP), type = "mu")
> summary(kpss.DRUSP)

#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 4 lags.

Value of test-statistic is: 0.0221

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

The test statistic (0.0221) is smaller than the 10% critical value (0.347), so H_0 is maintained at the 10% level, implying that the first difference of RUSP might be stationary.

All things considered, RCP, TXP, and RUSP prove to be stationary, i.e., integrated of order zero, I(0).

c) In the third stage of the project, the student intends to model *RCP*, *TXP*, *RUSP* with *VAR*. The relevant *R* commands and printouts are shown on the next page.

Briefly evaluate these printouts (use 5% significance level), mentioning the hypotheses, the statistical decisions and the conclusions. Based on these printouts, what is the order of your preferred *VAR* model? Explain your choice.

```
> library(vars)
> data = cbind(RCP, TXP, RUSP)
> VARselect(data, lag.max = 6, type = "const")
$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
     5      4      1      5

$criteria
                1         2         3         4         5         6
AIC(n)   4.074044  4.025801  3.943636  3.524186  3.487362  3.523562
HQ(n)    4.175075  4.202606  4.196215  3.852538  3.891488  4.003462
SC(n)    4.322674  4.460904  4.565212  4.332234  4.481883  4.704556
FPE(n) 58.796793 56.038339 51.641333 33.977403 32.792010 34.066133
```

```
> var1 = VAR(data, p = 1, type = "const")
> serial.test(var1, lags.bg = 5, type = "BG")

        Breusch-Godfrey LM test

data:  Residuals of VAR object var1
Chi-squared = 134.57, df = 45, p-value = 7.317e-11

> var2 = VAR(data, p = 2, type = "const")
> serial.test(var2, lags.bg = 5, type = "BG")

        Breusch-Godfrey LM test

data:  Residuals of VAR object var2
Chi-squared = 134.09, df = 45, p-value = 8.62e-11

> var3 = VAR(data, p = 3, type = "const")
> serial.test(var3, lags.bg = 5, type = "BG")

        Breusch-Godfrey LM test

data:  Residuals of VAR object var3
Chi-squared = 117.03, df = 45, p-value = 2.502e-08

> var4 = VAR(data, p = 4, type = "const")
> serial.test(var4, lags.bg = 5, type = "BG")

        Breusch-Godfrey LM test

data:  Residuals of VAR object var4
Chi-squared = 66.465, df = 45, p-value = 0.02036

> var5 = VAR(data, p = 5, type = "const")
> serial.test(var5, lags.bg = 5, type = "BG")

        Breusch-Godfrey LM test

data:  Residuals of VAR object var5
Chi-squared = 61.401, df = 45, p-value = 0.0523
```

The 4 model selection criteria selected three different lag lengths: SC one lag, HQ four lags, and AIC and FPE five lags.

The BG LM test was performed on the residuals of VAR models of order 1 to 5. In each test the null hypothesis is that the error variable has no autocorrelation of order 1 to 5, while the alternative hypothesis is that the error variable has some autocorrelation of order 1 to 5.

The p-values of the first four tests (7.317 / 10^{11}, 8.62 / 10^{11}, 2.502 / 10^{8}, 0.02036) are smaller than 0.05, but the p-value of the fifth test (0.0523) is larger than 0.05. Hence, the null hypothesis is rejected for the VAR models of orders 1 to 4, but maintained for the VAR(5) model. This means that the VAR(1), VAR(2), VAR(3) and VAR(4) models suffer from autocorrelation, but the error variable of the VAR(5) model does not have autocorrelation of orders 1 to 5.

For this reason, VAR(5) is the preferred specification.

d) Finally, the student performed Granger causality tests. The relevant *R* commands and printout are shown below:

```
> granger_causality(var5)

Granger Causality Test (Multivariate)

F test and Wald χ² test based on VAR(5) model:
────────────────────────────────────────────────────────
                    F df1 df2     p      Chisq df     p
────────────────────────────────────────────────────────
 -----------
 RCP  <= TXP     1.51   5 128  .190       7.57  5  .182
 RCP  <= RUSP    1.54   5 128  .182       7.70  5  .174
 RCP  <= ALL     1.82  10 128  .064 .    18.18 10  .052 .
 -----------
 TXP  <= RCP     3.71   5 128  .004 **   18.55  5  .002 **
 TXP  <= RUSP    0.76   5 128  .577       3.82  5  .576
 TXP  <= ALL     2.40  10 128  .012 *    24.04 10  .007 **
 -----------
 RUSP  <= RCP    2.63   5 128  .027 *    13.15  5  .022 *
 RUSP  <= TXP    1.81   5 128  .115       9.06  5  .107
 RUSP  <= ALL    2.60  10 128  .007 **   26.05 10  .004 **
────────────────────────────────────────────────────────
```

You have three tasks.

i. Evaluate the *Chisq* test results at the 5% significance level. For each of these tests report your statistical decision and conclusion with every numerical value from the printout that your answers rely on.

There are three sets of Granger causality tests, one set for each variable. In each test the null hypothesis is 'no Granger causality' and the alternative hypothesis is 'Granger causality'. In each set the first two tests are about *causality from one of the other two variables on the right side to the variable on the left side, while the third test is about joint causality from the two variables on the right side to the variable on the left side.*

The p-values in the first set are 0.182, 0.174 and 0.052. They are all above 0.05, so at the 5% significance level all three null hypotheses are maintained. This means that there is no evidence for Granger causality from TXP to RCP, from RUSP to RCP, and from (TXP, RUSP) to RCP.

The p-values in the second set are 0.002, 0.576 and 0.007. The second is above 0.05, but the first and the third are below 0.05. Hence, at the 5% significance level the first and the third null hypotheses are rejected and the second is maintained. This means that there is evidence for Granger causality running from RCP to TXP and from (RCP, RUSP) to TXP, but there is no evidence for Granger causality from RUSP to TXP.

The p-values in the third set are 0.022, 0.107 and 0.004. The second is above 0.05, but the first and the third are below 0.05. Hence, at the 5% significance level the first and the third null hypotheses are rejected and the second is maintained. This means that there is evidence for Granger causality running from RCP to RUSP and from (RCP, TXP) to RUSP, but there is no evidence for Granger causality from TXP to RUSP.

ii. Based on your answers in part (i), which variables prove to be endogenous in the *VAR*(5) model at the 5% significance level?

In a multivariate system, like this VAR5) model, a variable is considered an endogenous variable if the other variables jointly Granger cause it. Hence, based on the answers in part (i), only TXP and RUSP prove to be endogenous variables at the 5% significance level.

iii. Based on your answers in part (i), do you think that the student's project supports the two popular economic wisdoms about the economy of Texas USA, namely that "As the oil patch goes, so goes the Texas economy" and "Where the national economy goes, Texas economy need not follow"? Briefly explain your answers.

Given the three variables, the first wisdom suggests that the percentage quarterly change of the real price of crude oil (RCP) Granger causes the percentage quarterly change of the total number of nonfarm employment in the state of Texas (TXP). As discussed above in part (i), this wisdom is supported by the results.

The second wisdom suggests that the percentage quarterly change of the total number of nonfarm employment in the state of Texas (TXP) is not Granger caused by the percentage quarterly change of the total number of nonfarm employment in the other states of the US (RUSP). As discussed above in part (i), this wisdom is supported by the results.

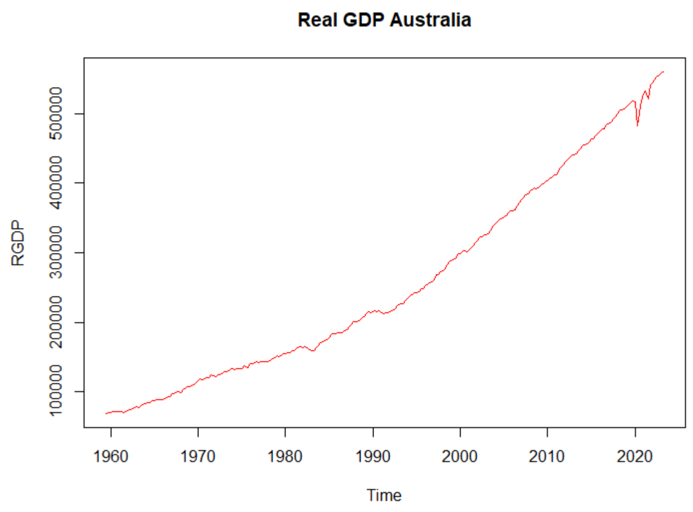
Question 4   (marks =)

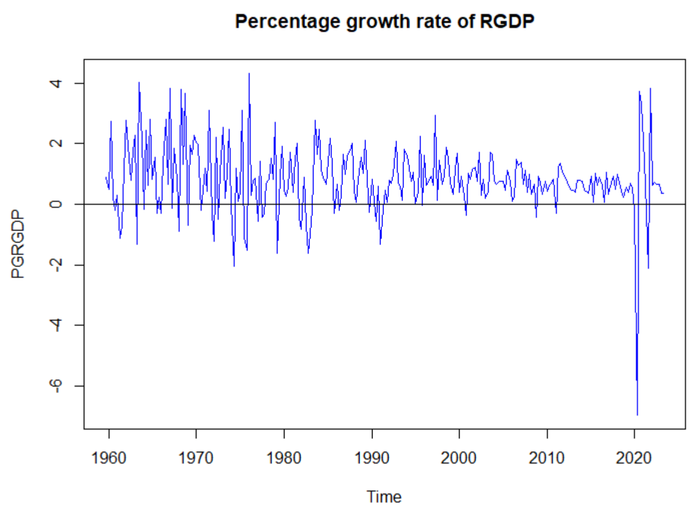
The Great Moderation refers to a period of unusually stable macroeconomic activity in the US and other advanced economies from the early 1980s to the financial crisis in 2007. It was characterised by a decrease in the volatility of the percentage growth rate of real GDP given by

$$PGRGDP_t = 100 \times \Delta \ln RGDP_t$$

where $RGDP_t$ is real GDP at time t.

The following graphs contain plots of quarterly data on $RGDP_t$ and $PGRGDP_t$ in Australia from the 3rd quarter 1959 to the 2nd quarter of 2023:





a) Based on the graphs above, briefly discuss the individual and combined effects of the logarithmic function and the first difference operator on the statistical properties of the *RGDP* series.

In general, the effect of the logarithmic function is twofold. First, it transforms an exponentially changing process into a linearly changing process, and second, it can stabilize volatility over time by dampening large movements relative to small movements.

The first difference operator can convert an I(1) process to an I(0) process.

The combination of these two filters converted the seemingly nonstationary level of real GDP into a series that looks stationary in the mean at least.

b) Consider the following model of $PGRGDP_t$:

$$PGRGDP_t = \alpha + \beta DGM_t + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

where ε_t is a white noise error term, DGM_t is a dummy variable defined as 1 for 1985 Q2 to 2019 Q4 and 0 otherwise, and (α, β, θ_1) are unknown parameters. The *R* printout from estimating this model is below:

```
Series: PGRGDP
Regression with ARIMA(0,0,1) errors

Coefficients:
          ma1  intercept     xreg
      -0.0896     0.8831  -0.1104
s.e.   0.0645     0.0981   0.1324

sigma^2 = 1.346:  log likelihood = -398.23
AIC=804.47   AICc=804.63   BIC=818.63
```

You have three tasks.

i. Write out the estimated equation using the proper notations of the variables.

The sample regression equation is

$$\widehat{PGRGDP}_t = 0.8831 - 0.1104 DGM_t + e_t - 0.0896 e_{t-1}$$

ii. Suppose there is a one standard deviation shock in time *t*, where *t* is between 1985 Q2 and 2019 Q4, inclusively. What are the effects of this shock on *PGRGDP* in times *t*, *t* +1 and *t* +2?

From the printout the sample variance of the error term is 1.346, so the supposed shock in time t is

$$\sqrt{1.346} = 1.160$$

The effect of this shock on $PGRGDP_t$ *is the shock itself, i.e., 1.160.*

To find the effect of this shock on $PGRGDP_{t+1}$, we need to update the sample regression equation to time t+1:

$$\widehat{PGRGDP}_{t+1} = 0.8831 - 0.1104DGM_{t+1} + e_{t+1} - 0.0896e_t$$

This shows that the effect is

$$-0.0896e_t = -0.0896 \times 1.160 = -0.1039$$

Likewise,

$$\widehat{PGRGDP}_{t+2} = 0.8831 - 0.1104DGM_{t+2} + e_{t+2} - 0.0896e_{t+1}$$

However, the shock in time t has no direct impact on the error variable in times t+ 1 and t+2, so its effect on $PGRGDP_{t+2}$ is zero.

iii. The residual of the estimated model in 2023 Q2 is about -0.5759%. Generate ex ante forecasts of $PGRGDP_t$ for the next three quarters.

The required ex ante forecasts are derived by writing out the sample regression equation and taking conditional expectations based on information at time T: 2023 Q2. Note that DGM_t is zero from 2020 Q1.

$$\begin{aligned} PGRGDP_{T+1} &= \alpha + \beta DGM_{T+1} + \varepsilon_{T+1} + \theta_1 \varepsilon_T \\ E_T(PGRGDP_{T+1}) &= \hat{\alpha} + \hat{\beta} DGM_{T+1} + E_T(\varepsilon_{T+1}) + \hat{\theta}_1 E_T(\varepsilon_T) \\ &= 0.8831 - 0.0896e_T = 0.8831 - 0.0896 \times (-0.5759) = 0.9347 \end{aligned}$$

Similarly,

$$\begin{aligned} PGRGDP_{T+2} &= \alpha + \beta DGM_{T+2} + \varepsilon_{T+2} + \theta_1 \varepsilon_{T+1} \\ E_T(PGRGDP_{T+2}) &= \hat{\alpha} + \hat{\beta} DGM_{T+2} + E_T(\varepsilon_{T+2}) + \hat{\theta}_1 E_T(\varepsilon_{T+1}) = 0.8831 \end{aligned}$$

and

$$\begin{aligned} PGRGDP_{T+3} &= \alpha + \beta DGM_{T+3} + \varepsilon_{T+3} + \theta_1 \varepsilon_{T+2} \\ E_T(PGRGDP_{T+3}) &= \hat{\alpha} + \hat{\beta} DGM_{T+3} + E_T(\varepsilon_{T+3}) + \hat{\theta}_1 E_T(\varepsilon_{T+2}) = 0.8831 \end{aligned}$$

End of questions