Econometrics – 2022-2023. Midterm exam. 2022 November 1.
Part 2. Free Response Questions (1 hour 30 minutes)
SUGGESTED SOLUTIONS

SECTION A Answer **ALL** questions from this section (questions **1-2**).

Question 1. (25 points) The following regressions were fitted using 37 annual observations for the period 1985-2021 for a certain country:

$$\widehat{\log L} = \underset{(0.224)}{10.984} - \underset{(0.024)}{0.093}\log Y \qquad R^2 = 0.30, \qquad (1)$$

$$\widehat{\log L} = \underset{(0.004)}{8.130} - \underset{(0.00014)}{0.00063}T \qquad (2)$$

$$\widehat{\log L} = \underset{(1.180)}{7.224} + \underset{(0.131)}{0.319}\log Y - \underset{(0.00073)}{0.00251}T \qquad R^2 = 0.48, \quad (3)$$

where: L – total employed labour force, measured in thousands,
Y – gross domestic product, measured in $ billion at constant 2020 prices, and
T – time, running from 1 to 37. (*The value of* R^2 *in equation* (2) *is not provided*)

(a) □ Give a careful interpretation of each equation in turn (slope coefficients in (1) and (3) and all coefficients in (2) including intercept).

Coefficient -0.093 in Equation (1) is the Y elasticity of L: a 1% increase in Y causes a 0.093% drop in L.
This Y elasticity of L in (3) is now positive 0.319 assuming T fixed.

In equation (2), a factor of -0.00063 indicates that L falls by -0.063% each year.
The drop of L in equation (3) is -0.251% each year assuming Y constant.

Intercept in equation (2) can be interpreted in the following way: the estimate of labor force for 1984 (corresponding zero T) is $e^{8.130} = 3395$ thousands (about 3.4 millions of people).

□ Are the coefficients of the variable $\log Y$ significant in equation (1) and in equation (3)?

Equation (1): $t_{\log Y} = -\frac{0.093}{0.024} = -3.875$, $|-3.875| > t_{crit}(1\%,\ 35) = 2.72$ – significant at the 1% level. Equation (3): $t_{\log Y} = \frac{0.319}{0.131} = -2.435 > t_{crit}(5\%,\ 33) = 2.032$, – significant at the 5% level.

□ Check with suitable tests whether equations (1) and (3) are significant in general?

Equation (1): It is enough to say that in the simple regression model the equation is significant if the slope coefficient is significant as it was shown earlier.
OR
$F = (t_{\log Y})^2 = (-3.875)^2 = 15$ or $F = \frac{0.3/1}{(1-0.3)/35} = 15$ or $F = (-3.875)^2 = 15$ while $F_{crit}(1\%, 1, 35) = 7.42$ - significant at the 1% level.
Equation (3): $F = \frac{0.48/2}{(1-0.48)/34} = 15.7$ while $F_{crit}(1\%, 2, 34) = 5.27$ – significant at the 1% level.

(b) □ Equation (1) can be viewed as a restricted version of equation (3). Indicate this restriction and run appropriate test.

Equation (3): $\log L = \beta_1 + \beta_2 \log Y + \beta_3 T + u$
equation (1): $\log L = \beta_1 + \beta_2 \log Y + u$
so the restriction is $H_0: \beta_3 = 0$ $(H_A: \beta_3 \neq 0)$. To test it we need to compare R^2 for two equations
$F = \frac{(0.48-0.3)/1}{(1-048)/34} = 11.77 > F_{crit}(1\%, 1, 34) = 7.42$ – significant at 1% level.
Alternative solution $t_T = \frac{-0{,}00251}{0{,}00073} \approx -3{,}43 \Rightarrow (3{,}43)^2 \approx 11{,}8$ with the same result.

□ Equation (2) is also a restricted version of equation (3). Indicate the restriction and run appropriate test(s).

The same logic is used for comparison (2) and (3), but first one should restore R^2 for (2):
$t_T = -\frac{0.00063}{0.00013} = -4.846$, so $F = (t_T)^2 = 23.485$, so solving $\frac{R^2/1}{(1-R^2)/35} = 23.485$ find $R^2 = 0.40$.
Now $F = \frac{(0.48-0.4)/1}{(1-048)/34} = 5.23 > F_{crit}(5\%, 1, 34) = 4.12$ – significant only at 5% level.
OR
The same result can be obtained more simply: we know that the value of F for comparing equations that differ by one variable is equal to the square of the t-statistic for this variable in the equation without restriction: from (3) $t_T = -\frac{0.00251}{0.00110} = -2.28$, $F = (t_T)^2 = 5.20$, the same conclusion.

□ Compare the coefficients for the same variables in equations (1), (2) and (3). Why do they differ in size, and some even sign? Suggest an explanation for the observed phenomena. Which equation would you recommend to choose for the further analysis as more credible? Explain.

Solution: Both tests for comparing equations showed a significant advantage of multiple regression (3) over simple regressions (1) and (2).
According to equation (3), $\log L$ is affected by two factors: $\log Y$ (positive influence) and T (negative influence). Since the factor T is not represented in equation (1), its negative effect is attributed to the factor $\log L$ present in the equation. As a result, the coefficient at the variable $\log Y$ ceases to express the real influence of this factor, and even becomes negative.
From this observation it is clear that simple regression, even with the significance of its coefficients, can erroneously describe the relationship of economic indicators.

Question 2. (25 points)
A student examines the dependence of public expenditure on private education ED_t (in million euros) in a small developed country on a time variable $TIME_t$ equal to 1 in 1997, and so on till 22 in 2018 (total 22 observations).

$$\widehat{ED}_t = 5.53 + 0.38 \cdot TIME_t \quad R^2 = 0.96, \qquad (1)$$

(a) □ Help the student to interpret the coefficients of the equation. Show that if calendar years $DATE_t$ are used instead of $TIME_t$, then the slope and R^2 remain the same, while the intercept changes.
A slope of 0.38 shows that spending on education increases by an average of 380,000 euros each year. Intercept 5.53 gives an estimate of spending on education in 1996 (the year preceding the first year in the sample).
When passing to the new variable $DATE_t$, the slope coefficient $\hat{\beta}_2$ of the regression $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X$ will not change since the new variable is related to the old by formula $Z = X + 1996$, so
$\hat{\beta}_2 = \frac{\widehat{\mathrm{Cov}}(X+1996,Y)}{\widehat{\mathrm{Var}}(X+1996)} = \frac{\widehat{\mathrm{Cov}}(X,Y)}{\widehat{\mathrm{Var}}(X)} = 0.38$. The intercept of the new regression is calculated as
$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2\overline{Z} = \overline{Y} - \hat{\beta}_2(\overline{X} + 1996) = (\overline{Y} - \hat{\beta}_2\overline{X}) - \hat{\beta}_2 \cdot 1996 = 5.53 - 0.38 \cdot 1996 = -752.95$
and obviously has no meaningful interpretation.
For the new regression, R^2 remains the same $R^2 = 0.96$, since $R^2 = 1 - \frac{SSR}{SST}$, where SST does not change, since it depends only on Y, and SSR is the sum of the squares of the residuals, where each of the residuals also does not change under a linear transformation

$$\hat{Y}_t - \hat{\beta}_1 - \hat{\beta}_2 Z_t = \hat{Y} - (5.53 - 0.38 \cdot 1996) - 0.38(X_t + 1996) = \hat{Y}_t - 5.53 - 0.38X_t$$

So the new equation

$$\widehat{ED}_t = -752.95 + 0.38 \cdot DATE_t \quad R^2 = 0.96,$$

□ Using Equation (1), help the student to obtain education expentidure forecast for 2019-2021 years. The actual amount of spending on education in these years was: in 2019 – 13.7, in 2020 – 13.6, in 2021 – 13.7. What conclusions about the quality of the forecast can be drawn from here?
Substituting 2019, 2020 and 2021 into equation (1) obtain forecasted values of ED_t and put them in a table

Actual	Forecast	Deviation	Deviation %
13.70000	14.34805	0.64805	4.730292
13.60000	14.73128	1.13128	8.318235
13.70000	15.11451	1.41451	10.32489

As one can see from the table, as you move away from the sample on which the regression is based, the deviations become larger, exceeding 10% in relative terms in 2021.

□ Indicate what factors affect the quality of the prediction (*you may use some formulas in your answer, but their proof is not required here*).
Let X^* be arbitrary value of explanatory variable X and $Y^* = \beta_1 + \beta_2 X^* + u^*$. We can get predicted value $\hat{Y}^* = \hat{\beta}_1 + \hat{\beta}_2 X^*$. The variance of the predicted value of the regression is equal to

$$\mathrm{Var}(\hat{Y}^*) = \left(\frac{1}{n} + \frac{(X^* - \overline{X})^2}{\sum(X_i - \bar{X})^2}\right)\sigma_u^2$$

Another approach can be based on the prediction error (PE)

$$PE = Y^* - \hat{Y}^* = (\beta_1 + \beta_2 X^* + u^*) - (b_1 + b_2 X^*)$$

Variance of the prediction error is equal to

$$\mathrm{Var}(PE) = \left(1 + \frac{1}{n} + \frac{(X^* - \overline{X})^2}{\sum(X_i - \overline{X})^2}\right)\sigma_u^2$$

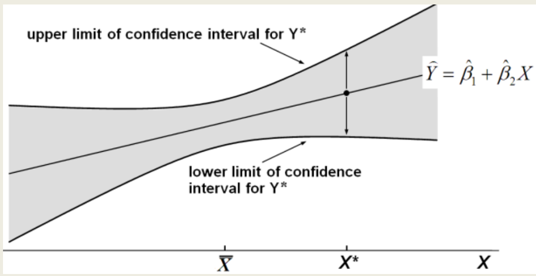
or standard error

$$s.e.(PE) = \sqrt{\left(1 + \frac{1}{n} + \frac{(X^* - \overline{X})^2}{\sum(X_i - \overline{X})^2}\right)\sigma_u^2}$$

As can be seen from these formulas, three main factors influence the forecast error:

1. variance of a random regression term σ_u^2 (direct dependence);
2. number of observations in the sample (inverse relationship)
3. deviation of the considered value of the independent variable X^* from its mean $\overline{X}$ (direct dependence).

The last dependence can also be expressed on the graph, which shows the change in the value of the confidence interval for the predicted value depending on the deviation $X^* - \overline{X}$.



(b) Let X^* be arbitrary value of explanatory variable X and $Y^* = \beta_1 + \beta_2 X^* + u^*$. Then predicted value is $\hat{Y}^* = \hat{\beta}_1 + \hat{\beta}_2 X^*$.

□ Show that $\mathrm{Cov}(Y^*, \hat{\beta}_2) = 0$

As it is known $\hat{\beta}_2 = \sum a_i Y_i$, where $a_i = \frac{(X_i - \overline{X})}{\sum(X_i - \overline{X})^2}$. So

$$\mathrm{Cov}(Y^*, \hat{\beta}_2) = \mathrm{Cov}(\beta_1 + \beta_2 X^* + u^*, \beta_2 + \sum a_i u_i) = \mathrm{Cov}(u^*, \sum a_i u_i) = \sum a_i \mathrm{Cov}(u^*, u_i) = 0$$

as $\mathrm{Cov}(u^*, u_i) = 0$ if $u^* \neq u_i$.

□ Show that $\mathrm{Cov}(Y^*, \hat{\beta}_1) = 0$

As it is known $\boldsymbol{b_1} = \boldsymbol{\beta_1} + \sum_{i=1}^{n} \boldsymbol{c_i u_i}$, where $\boldsymbol{c_1} = \frac{\mathbf{1}}{\boldsymbol{n}} - \boldsymbol{a_i \overline{X}}$ and a_i are defined as $\boldsymbol{a_i} = \frac{\boldsymbol{X_i - \bar{X}}}{\sum_{j=1}^{n}(\boldsymbol{X_j - \bar{X}})^2}$.

$$\mathrm{Cov}(Y^*, \hat{\beta}_1) = \mathrm{Cov}(\beta_1 + \beta_2 X^* + u^*, \beta_1 + \sum c_i u_i) = \mathrm{Cov}(u^*, \sum c_i u_i) = \sum c_i \mathrm{Cov}(u^*, u_i) = 0$$

Let us replace the predicted value Y^* in these formulas by the average value $\overline{Y}$.

□ Find what equals $\mathrm{Cov}(\overline{Y}, \hat{\beta}_2)$.

Using the same formula for decomposition of $\hat{\beta}_2 = \sum a_i Y_i$, one can get $\mathrm{Cov}(\overline{Y}, \hat{\beta}_2) = \mathrm{Cov}\left(\sum \frac{1}{n} Y_i, \sum a_i Y_i\right) = \sum \left(\frac{a_i}{n}\right) \mathrm{Var}(Y_i) + \sum_{i \neq j} \left(\frac{a_j}{n}\right) \mathrm{Cov}(Y_i, Y_j) = \frac{\sigma_u^2}{n} \sum a_i = 0$,
as $\mathrm{Cov}(Y_i, Y_j) = 0$ where $i \neq j$.

□ Find what equals $\mathrm{Cov}(\overline{Y}, \hat{\beta}_1)$

$$\mathrm{Cov}(\overline{Y}, \hat{\beta}_1) = \mathrm{Cov}(\overline{Y}, \hat{\beta}_1) = \mathrm{Cov}(\overline{Y}, \overline{Y} - \hat{\beta}_2 \overline{X}) = \mathrm{Cov}(\overline{Y}, \overline{Y}) + \overline{X}\,\mathrm{Cov}(\overline{Y}, \hat{\beta}_2) = \mathrm{Var}(\overline{Y}) + 0 = \frac{\sigma_u^2}{n}$$

Question 3. (25 points) At the seminar, the teacher asked the student to write down the formula from the lecture for the OLS estimation of the coefficient β_2 in the regression model:

$$Y_i = \beta_1 + \beta_2 X_i + u_i,\ i = 1,\ \dots,\ n,$$

The errors u_i are assumed to be independent random variables with zero mean and the regressor X_i is non-stochastic.

The student forgot the formula and wrote down the following expression

$$\hat{\beta}_2^{(1)} = \frac{\sum(X_i - \overline{X})Y_i}{\sum(X_i - \overline{X})X_i}$$

(a) □ Is this the formula for OLS estimator $\hat{\beta}_2^{OLS}$? What properties does the estimator $\hat{\beta}_2^{OLS}$ have? Give definitions and briefly explain the meaning of each of the mentioned properties.
(*You are not asked to derive $\hat{\beta}_2^{OLS}$ and prove its propernties – no points will be given for this*)

Formula for OLS estimator in simple linear regression is $\hat{\beta}_2^{OLS} = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2}$. OLS estimator is unbiased and efficient (Gauss-Markov theorem). Unbiasedness means $E(\hat{\beta}_2) = \beta_2$. It means that the expected value of the estimator is the true parameter; that is, we are correct on average in repeated samples. This ensures that we will not make systematic errors when estimating β.
The efficient estimator has the minimum variance among all linear unbiased estimators.

□ The teacher mentioned that the student's formula can also serve as a good estimator for β_2. Show that estimator $\hat{\beta}_2^{(1)}$ is unbiased.

$$\hat{\beta}_2^{OLS} = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2} = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})(X_i - \overline{X})} = \frac{\sum(X_i - \overline{X})Y_i - \overline{Y}\sum(X_i - \overline{X})}{\sum(X_i - \overline{X})X_i - \overline{X}\sum(X_i - \overline{X})} = \hat{\beta}_2^{(1)}$$

as $\sum(X_i - \overline{X}) = \sum X_i - \sum \overline{X} = n\overline{X} - n\overline{X} = 0$
As we know the OLS estimator $\hat{\beta}_2^{OLS}$ is unbiased and so $\hat{\beta}_2^{(1)}$ /

□ Is estimator $\hat{\beta}_2^{(1)}$ efficient?

As shown earlier $\hat{\beta}_2^{(1)} = \hat{\beta}_2^{OLS}$ and as we know $\hat{\beta}_2^{OLS}$ is efficient.

(b) □ Let Z_i be some non-stochastic variable different from X_i and Y_i. Consider the following estimator

$$\hat{\beta}_2^{*} = \frac{\sum(Z_i - \overline{Z})Y_i}{\sum(Z_i - \overline{Z})X_i}$$

where $\overline{Z} = \frac{1}{n}\sum Z_i$. Demonstrate that $\hat{\beta}_2^{*}$ is unbiased estimator for β_2.

When plugging in the true model ($Y_i = \beta_1 + \beta_2 X_i + u_i$), we can then write:

$$\hat{\beta}_2^{*} = \frac{\sum(Z_i - \overline{Z})Y_i}{\sum(Z_i - \overline{Z})X_i} = \frac{\sum(Z_i - \overline{Z})(\beta_1 + \beta_2 X_i + u_i)}{\sum(Z_i - \overline{Z})X_i} =$$

$$= \beta_1 \frac{\sum(Z_i - \overline{Z})}{\sum(Z_i - \overline{Z})X_i} + \beta_2 \frac{\sum(Z_i - \overline{Z})X_i}{\sum(Z_i - \overline{Z})X_i} + \frac{\sum(Z_i - \overline{Z})u_i}{\sum(Z_i - \overline{Z})X_i} = \beta_2 + \frac{\sum(Z_i - \overline{Z})u_i}{\sum(Z_i - \overline{Z})X_i}$$

as $\sum(Z_i - \overline{Z}) = \sum Z_i - \sum \overline{Z} = n\overline{Z} - n\overline{Z} = 0$.
So

$$\hat{\beta}_2^{*} = \beta_2 + \frac{\sum(Z_i - \overline{Z})u_i}{\sum(Z_i - \overline{Z})X_i}$$

And so taking expectations:

$$E(\hat{\beta}_2^*) = \beta_2 + \frac{\sum(Z_i - \bar{Z})\, E(u_i)}{\sum(Z_i - \bar{Z})X_i} = \beta_2$$

as X and so Z are assumed to be non-stochastic. Unbiasedness follows as $E(u_i) = 0$.

(c) □ Now consider another estimator

$$\hat{\beta}_2^{**} = \frac{\sum(Z_i - \bar{Z})Y_i}{\sum(Z_i - \bar{Z})Z_i}$$

Is this estimator also unbiased estimator for β_2?.

For this estimator, when plugging in the true model, ($Y_i = \beta_1 + \beta_2 X_i + u_i$), we obtain:

$$\hat{\beta}_2^{**} = \frac{\sum(Z_i - \bar{Z})Y_i}{\sum(Z_i - \bar{Z})Z_i} = \frac{\sum(Z_i - \bar{Z})(\beta_1 + \beta_2 X_i + u_i)}{\sum(Z_i - \bar{Z})Z_i} =$$

$$= \beta_1 \frac{\sum(Z_i - \bar{Z})}{\sum(Z_i - \bar{Z})Z_i} + \beta_2 \frac{\sum(Z_i - \bar{Z})X_i}{\sum(Z_i - \bar{Z})Z_i} + \frac{\sum(Z_i - \bar{Z})u_i}{\sum(Z_i - \bar{Z})Z_i} = \beta_2 \frac{\sum(Z_i - \bar{Z})X_i}{\sum(Z_i - \bar{Z})Z_i} + \frac{\sum(Z_i - \bar{Z})u_i}{\sum(Z_i - \bar{Z})X_i}$$

as $\sum(Z_i - \bar{Z}) = \sum Z_i - \sum \bar{Z} = n\bar{Z} - n\bar{Z} = 0$.

$$\hat{\beta}_2^{**} = \beta_2 \frac{\sum(Z_i - \bar{Z})X_i}{\sum(Z_i - \bar{Z})Z_i} + \frac{\sum(Z_i - \bar{Z})u_i}{\sum(Z_i - \bar{Z})Z_i}$$

and:

$$E(\hat{\beta}_2^{**}) = \beta_2 \frac{\sum(Z_i - \bar{Z})X_i}{\sum(Z_i - \bar{Z})Z_i} \neq \beta_2$$

using again the fact that Z is non-stochastic and $E(u_i) = 0$, hence $\hat{\beta}_2^{**}$ is biased.

□ Indicate the regression for which the estimate $\hat{\beta}_2^{**}$ serves as the OLS estimator of the slope coefficient.

The last expression is OLS estimator $\hat{\gamma}_2^{OLS}$ of the slope coefficient γ_2 in the regression $Y_i = \gamma_1 + \gamma_2 Z_i + u_i$, so $\hat{\beta}_2^{**} = \hat{\gamma}_2^{OLS}$.

Question 4. (25 points) The supervisor assigned three students A, B and C to investigate the dependence of the gross agricultural production $agro$ of a small developing country on gdp (both measured in billions of dollars) as well as population pop (measured in millions). Based on the same data for 47 years, each of them calculated his own equation (RSS = residual sum of squares):

$$\widehat{\log\frac{agro}{pop}} = -3.74 + 1.19\log\frac{gdp}{pop} \qquad R^2 = 0.9094,\quad SSR = 14.26$$
(A)

$$\widehat{\log agro} = -3.60 + 1.27\log gdp - 0.33\log pop \qquad R^2 = 0.9531,\quad SSR = 13.90$$
(B)

$$\widehat{\log\frac{agro}{pop}} = -3.60 + 1.27\log\frac{gdp}{pop} - 0.06\log pop \qquad R^2 = 0.9117,\quad SSR = 13.90$$
(C)

(a) □ Explain why coefficient of variable $\log\frac{gdp}{pop}$ in equation (C) is equal to the coefficient of variable $\log gdp$ in equation (B).

In equation (C) the coefficient of $\log\frac{gdp}{pop}$ formally shows the **net effect** of GDP per capita, but since the population size is fixed, this effect is achieved solely by GDP, so that their coefficients should be equal.

OR

Assuming pop constant it can be shown that elasticity for eq.C is the same as for eq.B

$$\beta_2(\text{eq.C}) = \frac{d\left(\frac{agro}{pop}\right)}{d\left(\frac{gdp}{pop}\right)}\cdot\frac{\left(\frac{gdp}{pop}\right)}{\left(\frac{agro}{pop}\right)} = \frac{\frac{1}{pop}d(agro)}{\frac{1}{pop}d(gdp)}\cdot\frac{gdp}{agro} = \frac{d(agro)}{d(gdp)}\cdot\frac{gdp}{agro} = \beta_2(\text{eq.B}).$$

OR

$\log\frac{agro}{pop} = \beta_1 + \beta_2\log\frac{gdp}{pop} + \beta_3\log pop + u \Leftrightarrow \log agro = \beta_1 + \beta_2\log gdp + (\beta_3 - \beta_2 - 1)\log pop + u$

□ Explain why equations (B) and (C) have the same intercept, the same SSR, but different R^2?

$$\log\frac{agro}{pop} = \beta_1 + \beta_2\log\frac{gdp}{pop} + \beta_3\log pop + u$$
$$\log agro - \log pop = \beta_1 + \beta_2\log gdp - \beta_2\log pop + \beta_3\log pop + u$$
$$\log agro = \beta_1 + \beta_2\log gdp\log pop + (\beta_3 - \beta_2 + 1)\log pop + u$$

So both equations actually estimate the dependence of $\log agro$ on $\log gdp$ and $\log pop$. Therefore, they must have the same intercept, and the same residual sum of squares SSR. However, they have different dependent variables, and hence different SST values, so their R^2 should be different.

(b) □ Demonstrate that equation (A) is a restricted version of equation (C), stating the restriction.

Comparing the the first specification

$$\log\frac{agro}{pop} = \beta_1 + \beta_2\log\frac{gdp}{pop} + u.$$

with the second one

$$\log\frac{agro}{pop} = \beta_1 + \beta_2\log\frac{gdp}{pop} + \beta_3\log pop + u.$$

we can see that restriction is $\beta_3 = 0$.

□ Student A tested this restriction using R^2, and student C used SSR for this purpose. Which of them is right? How will you proceed the appropriate test(s) and what are your conclusions?

The null hypothesis is $\beta_3 = 0$.
Both tests are correct and give the same result.
$F = \frac{(14.26-13.90)/1}{13.90/44} = 1.14$, and $F = \frac{(0.9117-0.9094)/1}{(1-0.9117)/44} = 1.14$. The critical value of $F(1,44)$ at the 5 percent significance level is about 4.08. Hence we do not reject the restriction.

(c) □ Demonstrate that (A) is a restricted version of equation (B), stating the restriction.

Write the first specification

$$\log \frac{agro}{pop} = \beta_1 + \beta_2 \log \frac{gdp}{pop} + u.$$

It may be rewritten as

$$\log agro - \log pop = \beta_1 + \beta_2 \log gdp - \beta_2 \log pop + u$$

and this in turn may be rewritten

$$\log agro = \beta_1 + \beta_2 \log gdp + (1-\beta_2) \log pop + u$$

This is a restricted version of the more general specification

$$\log agro = \beta_1 + \beta_2 \log gdp + \beta_3 \log pop + v$$

with the restriction $\beta_3 = 1 - \beta_2$. .

□ Student A tested this restriction using R^2, and student B used SSR for this purpose. Which of them is right? How will you proceed the appropriate test(s) and what are your conclusions?

The null hypothesis is $H_0 : \beta_3 = 1 - \beta_2$. Test on the base of SSR is always correct $F = \frac{(14.26-13.90)/1}{13.90/44} = 1.14$.
The critical value of $F(1,44)$ at the 5 percent significance level is about 4.08. Hence we do not reject the restriction. Student B is right.
It is impossible to use test based on R^2 here as dependent variables of equations (1) and (3) are different so their SST are different what makes their $R^2 = 1 - \frac{SSR}{SST}$ are incomparable.