Continuous Assessment in Applied Econometrics – 2H

Anderson Vil
CY Cergy Paris Université
2026 – 2027

Master 1 – Mock exam 2 – Solutions

Answer the questions in order. All answers must be justified. Calculators are allowed. Throughout, log denotes the natural logarithm. Unless stated otherwise, the assumptions of the classical linear model, including normally distributed errors, are taken to hold. Critical values are read from the tables at the end of the paper. When the number of degrees of freedom is not printed, use the closest lower row, and use the row ∞ above 120. The marking scheme is indicative.

Exercise 1: Days at sea and catches of a trawler fleet *(5 points)*

A fishermen's cooperative in Brittany records, for $N = 12$ trawlers, the number of days spent at sea in October 2026 (x_i) and the catch (y_i, in tonnes). For the model $y = X\beta + u$, where the columns of X are a constant and x, the cooperative's software prints:

$$X'X = \begin{pmatrix} 12 & 216 \\ 216 & 4{,}104 \end{pmatrix}, \qquad X'y = \begin{pmatrix} 360 \\ 6{,}804 \end{pmatrix}, \qquad y'y = 11{,}346.$$

1. Compute $(X'X)^{-1}$ and $\hat\beta = (X'X)^{-1}X'y$. Interpret $\hat\beta_1$. *[1.25]*

Solution. $\det(X'X) = 12 \times 4104 - 216^2 = 49248 - 46656 = 2592$; $(X'X)^{-1} = \frac{1}{2592}\begin{pmatrix} 4104 & -216 \\ -216 & 12 \end{pmatrix}$. $\hat\beta = \frac{1}{2592}\begin{pmatrix} 4104 \times 360 - 216 \times 6804 \\ -216 \times 360 + 12 \times 6804 \end{pmatrix} = \frac{1}{2592}\begin{pmatrix} 7776 \\ 3888 \end{pmatrix} = \begin{pmatrix} 3 \\ 1.5 \end{pmatrix}$. One more day at sea is associated with 1.5 more tonnes of catch.

2. Let $M = I - X(X'X)^{-1}X'$, so that the residual vector is $\hat u = My$. Using the fact that M is symmetric and idempotent, show that $\hat u'\hat u = y'y - \hat\beta' X'y$. Deduce the sum of squared residuals without computing any residual. *[1.25]*

Solution. $\hat u'\hat u = (My)'(My) = y'M'My = y'MMy = y'My$ (symmetry, then idempotence). $y'My = y'y - y'X(X'X)^{-1}X'y$ and, since $(X'X)^{-1}$ is symmetric, $y'X(X'X)^{-1} = \big((X'X)^{-1}X'y\big)' = \hat\beta'$. Hence $\hat u'\hat u = y'y - \hat\beta' X'y$. Numerically: $\hat\beta' X'y = 3 \times 360 + 1.5 \times 6804 = 11286$, so SSR $= 11346 - 11286 = 60$.

3. Compute $\hat\sigma^2$, the estimated variance matrix $\hat\sigma^2(X'X)^{-1}$, and the standard errors of $\hat\beta_0$ and $\hat\beta_1$. *[1]*

Solution. $\hat\sigma^2 = 60/(12-2) = 6$. $\hat\sigma^2(X'X)^{-1} = \frac{6}{2592}\begin{pmatrix} 4104 & -216 \\ -216 & 12 \end{pmatrix} = \begin{pmatrix} 9.5 & -0.5 \\ -0.5 & 0.0278 \end{pmatrix}$. $\mathrm{se}(\hat\beta_0) = \sqrt{9.5} = 3.08$, $\mathrm{se}(\hat\beta_1) = \sqrt{0.0278} = 0.167$.

4. A regional fisheries report states that one more day at sea yields 2 tonnes on average. Test this statement at the 5% level against a two-sided alternative. *[0.75]*

Solution. $H_0 : \beta_1 = 2$. $t = (1.5 - 2)/0.167 = -3$, $|t| > t_{0.975}(10) = 2.228$: rejected.

5. Suppose that all trawlers had spent exactly 18 days at sea. Write $X'X$ in that case and explain, in two or three lines, why β_1 could not be estimated. *[0.75]*

Solution. $\sum x_i = 216$, $\sum x_i^2 = 12 \times 18^2 = 3888$: $X'X = \begin{pmatrix} 12 & 216 \\ 216 & 3888 \end{pmatrix}$, whose determinant is $46656 - 46656 = 0$. The column x equals 18 times the constant ($Xc = 0$ with $c = (18, -1)'$): X is not of full column rank, $X'X$ is not invertible. Without variation in days at sea, the data cannot distinguish the effect of days from the intercept.

Typical mistakes. forgetting to divide by the determinant; writing $y'X(X'X)^{-1}X'y = \hat{\beta}X'y$ without the transpose (a 2×2 object instead of a scalar); using N instead of $N-2$; reading $\text{se}(\hat{\beta}_1)$ as 0.0278 (variance, not standard error).

Exercise 2: Who consumes electricity? ***(6 points)***

An energy regulator studies the annual electricity consumption of 800 French households. The dependent variable is the log of annual consumption (kWh). The regressors are the log of the household's annual disposable income (euros), the log of the dwelling's surface (m^2), *eheat*, a dummy equal to 1 if the dwelling has electric heating, and *occ*, the number of occupants. Column (4) reports the regression of $\log(surf)$ on a constant and $\log(inc)$.

Table 1 – Household electricity consumption, OLS estimates

	(1)	(2)	(3)	(4)
Dependent variable	$\log(kwh)$	$\log(kwh)$	$\log(kwh)$	$\log(surf)$
$\log(inc)$	0.349	0.079	0.056	0.377
	(0.043)	(0.044)	(0.037)	(0.025)
$\log(surf)$		0.716	0.736	
		(0.056)	(0.047)	
eheat			0.469	
			(0.030)	
occ			0.106	
			(0.010)	
Constant	4.799	4.486	4.154	0.437
	(0.450)	(0.410)	(0.344)	(0.261)
N	800	800	800	800
R^2	0.076	0.235	0.468	0.224

Notes: standard errors in parentheses.

We denote by $\tilde{\beta}_1$ the coefficient on $\log(inc)$ in column (1), by $\hat{\beta}_1$ and $\hat{\beta}_2$ the coefficients on $\log(inc)$ and $\log(surf)$ in column (2), and by $\tilde{\delta}_1$ the coefficient on $\log(inc)$ in column (4).

1. Write the estimated equation of column (2). Interpret the coefficients on $\log(inc)$ and on $\log(surf)$. *[1]*

Solution. $\widehat{\log(kwh)} = 4.486 + 0.079\log(inc) + 0.716\log(surf)$. For a given surface, a 1% higher income is associated with 0.08% higher consumption; for a given income, a 1% larger dwelling is associated with 0.72% higher consumption (elasticities).

2. In column (2), test at the 5% level the hypothesis that income has no effect on consumption for a given surface, against a two-sided alternative. Same question at the 10% level. *[0.75]*

Solution. $t = 0.079/0.044 = 1.80$. $1.80 < t_{0.975}(797) \approx 1.960$ (row ∞): not rejected at 5%. $1.80 > t_{0.95}(797) \approx 1.645$: rejected at 10%.

3. Using only the R^2 of column (2), test at the 5% level the joint significance of its regressors. *[0.75]*

Solution. $F = \dfrac{0.235/2}{(1-0.235)/(800-3)} = \dfrac{0.1175}{0.00096} = 122.4 > F_{0.95}(2,797) \approx 3.00$ (row ∞): rejected.

4. Test at the 5% level whether *eheat* and *occ* can be jointly excluded from column (3). Give the exact percentage effect of electric heating. *[1]*

Solution. $F = \dfrac{(0.468-0.235)/2}{(1-0.468)/(800-5)} = \dfrac{0.1165}{0.000669} = 174.1 > F_{0.95}(2,795) \approx 3.00$ (row ∞): rejected. Electric heating: $e^{0.469} - 1 = +59.8\%$ consumption, other things equal.

5. (a) Interpret $\tilde{\delta}_1$.

(b) Verify numerically the relation $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2\tilde{\delta}_1$.

(c) Is the income elasticity overestimated or underestimated in column (1)? By how much? Explain in one sentence.

[1.5]

Solution. (a) A 1% higher income is associated with a 0.38% larger dwelling.
(b) $0.079 + 0.716 \times 0.377 = 0.079 + 0.270 = 0.349 = \tilde{\beta}_1$.
(c) The bias $\hat{\beta}_2\tilde{\delta}_1 = +0.27$ is positive: column (1) overestimates the elasticity by 0.27, because richer households live in larger dwellings, which consume more.

6. On the basis of column (1), a newspaper writes: "a 10% rise in income raises a household's electricity consumption by 3.5%". Comment in two or three lines. *[1]*

Solution. Column (1) gives the total association, which mostly works through the size of the dwelling (0.27 out of 0.35); for a given dwelling, the association is only 0.08 (0.8% for +10%) and not significant at 5%. "Raises" is a causal claim: even column (3) may omit factors correlated with income (appliances, time spent at home), so neither number is a causal effect without further assumptions.

Typical mistakes. reading 0.079 as "7.9%"; using $N - 5$ degrees of freedom in question 3; taking column (1) as restricted model in question 4; wrong sign of the bias; treating 0.469 as a 46.9% effect without comment.

Exercise 3: Waste per inhabitant and the size of municipalities *(4 points)*

Two analysts of a regional waste agency use the same data on $N = 200$ municipalities. Analyst A estimates by OLS:

$$\widehat{\log(waste)} = \underset{(0.48)}{-1.43} + \underset{(0.05)}{1.08}\log(pop), \qquad R^2 = 0.702, \qquad \text{SSR} = 17.82,$$

where *waste* is the household waste collected (tonnes per year) and *pop* the number of inhabitants. Analyst B regresses $\log(waste/pop)$, the log of waste per inhabitant, on a constant and $\log(pop)$.

1. Without re-estimating, give the intercept, the slope, the standard error of the slope and the SSR of analyst B's regression. Justify each answer. *[1.25]*

Solution. $\log(waste/pop) = \log(waste) - \log(pop)$: subtracting the regressor from the dependent variable lowers the slope by one and leaves the intercept unchanged: intercept -1.43, slope $1.08 - 1 = 0.08$. The residuals are identical: $\log(waste/pop) - (-1.43) - 0.08\log(pop) = \log(waste) - (-1.43) - 1.08\log(pop)$. Hence SSR $= 17.82$, same $\hat{\sigma}^2$, and same regressor, so the same standard error 0.05.

2. Interpret the slope of analyst A and the slope of analyst B. *[0.75]*

Solution. A: a municipality with 1% more inhabitants collects 1.08% more waste. B: a municipality with 1% more inhabitants collects 0.08% more waste *per inhabitant*. The two statements are the same result.

3. Test at the 5% level the hypothesis that waste per inhabitant does not depend on the size of the municipality, against a two-sided alternative. To which hypothesis on analyst A's slope is it equivalent? *[0.75]*

Solution. $H_0 : \beta_2 = 0$ in B, equivalent to $H_0 : \alpha_2 = 1$ in A (waste proportional to population). $t = 0.08/0.05 = 1.6 < t_{0.975}(198) \approx 1.960$ (row ∞): not rejected.

4. Analyst B reports $R^2 = 0.013$. Analyst A concludes: "my model explains 70% of the variance and B's model almost nothing, so B's model is worse". Compute the total sum of squares of each regression and explain why A's conclusion is wrong. *[0.75]*

Solution. $\text{TSS}_A = 17.82/(1-0.702) = 59.8$, $\text{TSS}_B = 17.82/(1-0.013) = 18.05$. Both models leave exactly the same residuals; the R^2s differ only because the dependent variables, hence the variations to be explained, differ (municipality size explains most of the variation of total waste, little of the variation of waste per inhabitant). The two models fit equally well.

5. An intern proposes to test the hypothesis of question 3 with $F = \dfrac{(0.702-0.013)/1}{(1-0.702)/198}$. Explain in two or three lines why this statistic is meaningless. *[0.5]*

Solution. Model B is not the restricted model: it is model A rewritten, with the same SSR (17.82), so it imposes no restriction. Under H_0 the restricted model is $\log(waste/pop) = \alpha_0 + v$, whose SSR is $\text{TSS}_B = 18.05$. Moreover, the R^2 form of F requires the same dependent variable in both models (same TSS), which fails here. Correct statistic: $F = \dfrac{(18.05-17.82)/1}{17.82/198} = 2.56 = t^2 = 1.6^2$, equivalently $F = \dfrac{R_B^2}{(1-R_B^2)/198}$: same conclusion as question 3.

Typical mistakes. slope of B equal to $1.08/1$ or -0.08; changing the intercept by $\log(pop)$; claiming the standard error changes because R^2 changes; testing $\alpha_2 = 0$.

Exercise 4: How much do Parisians tip? ***(5 points)***

A restaurant chain in Paris records, for $N = 250$ tables, the bill (euros, between 15 and 180) and the tip left by the customers (euros). It estimates by OLS (standard errors in parentheses):

$$\widehat{tip} = \underset{(0.30)}{0.80} + \underset{(0.007)}{0.100}\, bill, \qquad R^2 = 0.45.$$

1. Interpret the slope and the intercept. *[0.75]*

Solution. Slope: each extra euro on the bill is associated with 10 cents of extra tip (marginal tip rate 10%). Intercept: predicted tip for a zero bill, 0.80 euro; outside the sample range, it is best read as the part of the tip that does not scale with the bill (a "round-up" coin).

2. Write the predicted average tip rate $\widehat{tip}/bill$ as a function of the bill, and compute it for bills of 20 and 100 euros. With the bill on the horizontal axis, sketch the marginal and the average tip rates, and compare them. *[1]*

Solution. $\widehat{tip}/bill = 0.10 + 0.80/bill$: 14% for 20 euros, 10.8% for 100 euros. The marginal rate is a horizontal line at 0.10; the average rate is decreasing and convex, always above it (positive intercept), and tends to 0.10 as the bill grows.

3. A guidebook states that "Parisian diners tip 12% of the bill". Write the two restrictions this statement imposes on β_0 and β_1, and test each of them separately at the 5% level against a two-sided alternative. *[1.25]*

Solution. The statement means $tip = 0.12\, bill$: $\beta_0 = 0$ and $\beta_1 = 0.12$. $t = 0.80/0.30 = 2.67$ and $t = (0.100-0.12)/0.007 = -2.86$; both exceed $t_{0.975}(248) \approx 1.960$ (row ∞) in absolute value: both restrictions are rejected. (A joint test is not required.)

4. The dispersion of tips increases with the bill. Which assumption fails? What are the consequences for the OLS estimates and for the tests of question 3? *[1]*

Solution. Homoskedasticity fails. OLS remains unbiased (under $E(u \mid bill) = 0$) but is no longer the most precise linear unbiased estimator, and the usual standard errors are wrong: the t statistics of question 3 do not follow a Student distribution, so the conclusions must be checked with robust standard errors.

5. Is 0.100 the causal effect of a higher bill on the tip? Give two factors contained in the error term and the plausible direction of the bias. *[1]*

Solution. Not necessarily. Examples: quality of service, celebrations (birthdays), customers' income, group size. Celebrations and richer customers raise both the bill and the generosity of the tip: positive correlation with the bill and positive effect on the tip, so the slope is biased upward. Large groups may tip proportionally less, which would bias it downward. Any coherent pair (factor, sign) gets the marks.

Typical mistakes. average rate $0.10 + 0.80$; testing $\beta_1 = 12$; one-sided tests; claiming heteroskedasticity biases the slope.