Continuous Assessment in Applied Econometrics – 2 hours

Anderson Vil
CY Cergy Paris Université
2026 – 2027

Detailed solutions – Mock exam 2

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.

Part A – Auditing the table *(4 points)*

A1. $10 \times 0.1366 = 1.37$ hours, which is correct. Then $t = 0.1366/0.0102 = 13.4 > t_{0.995}(2{,}394) \approx 2.576$, so the effect is indeed significant at 1% (two-sided test). Claim (i) is accurate.

Marks: 0.5 effect + 0.5 t and level = 1

A2. Column (4) is column (2) under a linear restriction, on the same sample and with the same dependent variable. Imposing a constraint cannot reduce the sum of squared residuals, so $\text{SSR}_4 \geq \text{SSR}_2$ and $R^2_4 \leq R^2_2$. This is the case here ($0.2539 \leq 0.2558$). But "fitting almost as well" does not validate the restriction. Only a test can do that (C3).

Marks: 0.5 inequality and justification + 0.5 consistency and limit of the argument = 1

A3. The new dummy equals $1 - x_2$, so it adds up with x_2 to the column of ones. With $c = (1, 0, -1, 0, 0, 0, -1)' \neq 0$ (constant, x_1, x_2, x_3, x_4, x_5, new dummy) we get $Xc = 0$, so $\text{rank}(X) < K$ and $c'X'Xc = 0$. $X'X$ is singular and the coefficients are not identified (dummy variable trap). The software must have dropped one dummy (or the constant). In column (2), with the constant kept, the coefficient on "Two or more children" is the gap in weekly hours between mothers of two or more children and mothers of one child, the reference category, other things being equal.

Marks: 0.5 diagnosis via $X'X$ + 0.5 what the software did and how to read the coefficient = 1

A4. $\hat{\sigma}^2(X'X)^{-1}$ is positive definite (the inverse of a positive definite matrix, since X has full column rank). Every 2×2 principal submatrix is then positive definite too, so its determinant is positive: $\widehat{\text{Cov}}(\hat{\beta}_j, \hat{\beta}_k)^2 < \widehat{\text{Var}}(\hat{\beta}_j)\widehat{\text{Var}}(\hat{\beta}_k)$. For $(\hat{\beta}_1, \hat{\beta}_3)$ we have $0.0102 \times 0.084 = 0.00086 < 0.0021$, an implied correlation of 2.45, which is impossible. The covariance of $(\hat{\beta}_4, \hat{\beta}_5)$ satisfies the bound ($0.192 \times 0.421 = 0.081 > 0.0502$), and so does the one in column (3). Plausible explanation: a misplaced decimal point (a value of the order of 0.00002 would be admissible).

Marks: 0.5 identification with the bound + 0.25 justification (positive definiteness) + 0.25 explanation = 1

Typical mistakes.

- Using 1.96 for a 1% test.
- In A2, believing that the ordering of R^2s depends on whether the restriction is true.
- In A3, speaking of "strong multicollinearity": the dependence is exact and $X'X$ cannot be inverted.
- In A4, declaring "impossible" a covariance that is merely large. The criterion is the bound $\text{s.e.}_j \times \text{s.e.}_k$.

Part B – Assessing the results *(5 points)*

B1. For $\Delta x_1 = 10$ and given education and income, the effect is $10\,\hat{\beta}_1 = 1.74$ hours per week for a mother of one child and $10(\hat{\beta}_1 + \hat{\beta}_6) = 10 \times 0.0906 = 0.91$ hours for a mother of two or more children. Childcare places raise labour supply, but much less when the family is larger. A place for the youngest child does not free the time spent on the older ones.

Marks: 0.5 first effect + 0.5 second effect + 0.5 interpretation = 1.5

B2. For given places, number of children, education and other income, an extra 1,000 euros of the partner's monthly income is associated with -1.255 hours of work per week (about 75 minutes less). This is an income effect, leisure being a normal good.

Marks: 1

B3. $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ and column (1) is the restricted model. Both models have the same h and the same sample and are nested, so

$$F = \frac{(0.2558 - 0.0580)/4}{(1 - 0.2558)/2{,}394} = \frac{0.04945}{0.0003109} = 159.1 > F_{0.95}(4,\, 2{,}394) \approx 2.37$$

and we reject H_0. The coefficient on places only moves from 0.1394 to 0.1366. By the omitted variable identity, this shift equals $\sum_k \hat{\beta}_k \hat{\delta}_k$. It is small because the supply of places is almost uncorrelated with the individual characteristics added. This is reassuring for these variables, but not for omitted municipal factors (see D1–D2).

Marks: 0.75 test + 0.75 reading of the stability = 1.5

B4. Under heteroskedasticity OLS remains unbiased, but $\hat{\sigma}^2 (X'X)^{-1}$ is wrong, so all the standard errors, t and F statistics of the paper are invalid. Claims (iii) and (iv) rest on tests whose statistics are not far from the critical values, and claim (ii) rests on the interaction, estimated mainly on the noisier group. They are the most fragile. Claim (i), with $t = 13.4$, will very probably survive. Standard errors valid under heteroskedasticity are needed.

Marks: 0.5 consequence + 0.5 sorting the claims = 1

Typical mistakes.

- Forgetting the interaction and answering $10\hat{\beta}_1$ for all mothers.
- Reading $\hat{\beta}_4$ as an effect "per euro".
- Claiming that heteroskedasticity biases the coefficients.

Part C – Income pooling ***(5 points)***

C1. $H_0 : \beta_4 = \beta_5$. In the unitary model, the household maximises a single utility function under a single budget constraint: only total non-labour income $x_4 + x_5$ enters the constraint, and its source is irrelevant.

Marks: 0.25 hypothesis + 0.25 justification = 0.5

C2. $\sqrt{0.192^2 + 0.421^2} = \sqrt{0.0369 + 0.1772} = \sqrt{0.2141} = 0.463$: the authors forgot the covariance. The correct calculation is

$$\widehat{\mathrm{Var}}(\hat{\beta}_4 - \hat{\beta}_5) = 0.2141 - 2 \times 0.0502 = 0.1137 \qquad \text{s.e.} = 0.337 \qquad t = -0.830/0.337 = -2.46$$

Since $|t| > t_{0.975}(2{,}394) \approx 1.960$, income pooling is rejected at the 5% level. The 95% confidence interval is $[-1.491,\ -0.169]$. The covariance is positive because x_4 and x_5 are negatively correlated (benefits fall when the partner's income rises). Forgetting it inflates the variance of the difference and reverses the conclusion.

Marks: 0.5 reconstruction of the error + 0.75 correct variance + 0.5 decision + 0.25 interval = 2

C3. Column (4) replaces x_4 and x_5 by their sum, which has a single coefficient, so $\beta_4 = \beta_5$ holds by construction. Both models have the same h and the same sample and are nested, with $Q = 1$

$$F = \frac{(R_2^2 - R_4^2)/1}{(1 - R_2^2)/2{,}394} = \frac{0.0019}{0.0003109} = 6.11$$

With a single restriction $F = t^2$, and we recover $(-2.46)^2 = 6.06$ up to rounding, not the authors' $(-1.79)^2 = 3.22$. Since $F > F_{0.95}(1,\, 2{,}394) \approx 3.84$, column (4) imposes a restriction rejected by the data, and the authors had the means to see it themselves.

Marks: 0.5 mechanism of column (4) + 0.5 calculation of F + 0.5 settling the disagreement through $F = t^2$ = 1.5

C4. $t = -1.741/0.594 = -2.93$. Table A2 has the regressors of column (2), so $K = 6$ and there are $N - 6$ degrees of freedom. We reject if $|t| > t_{0.975}(N-6)$. In Table G.2, $t_{0.975}(3) = 3.182 > -2.93$ (no rejection) but $t_{0.975}(4) = 2.776 < -2.93$ (rejection), and the critical value keeps decreasing as the degrees of freedom increase. The hypothesis is therefore rejected as soon as $N - 6 \geq 4$, that is $N \geq 10$. A regional subsample of a survey of 2,400 mothers almost certainly contains more observations. In practice the conclusion does not depend on N: income pooling is rejected in Île-de-France as well, and claim (iv) is wrong.

Marks: 0.5 degrees-of-freedom threshold + 0.5 conclusion = 1

Typical mistakes.

- Counting the covariance only once, or with the wrong sign.
- In C3, forgetting that $F = t^2$ only holds for $Q = 1$, the same sample and the same dependent variable.
- In C4, using 1.96 "because N is unknown".

Part D – What the referee asks for ***(6 points)***

D1. By Frisch–Waugh–Lovell, the coefficient of x_1 in the augmented regression is

$$\hat{\beta}_1^{(z)} = \frac{\sum_i \hat{r}_i h_i}{\sum_i \hat{r}_i^2} = \frac{113{,}478.5}{680{,}920} = 0.1667$$

For given children, education, income and commuting time, ten additional places are associated with 1.67 more hours of work per week.

Marks: 1 calculation + 0.5 interpretation = 1.5

D2. Omitted-variable identity, exact in the sample: $\tilde{\beta}_1 = \hat{\beta}_1^{(z)} + \hat{\beta}_z \hat{\delta}_1$, where $\tilde{\beta}_1 = 0.1366$ (column (2)) and $\hat{\delta}_1$ is the coefficient sought. Hence

$$\hat{\delta}_1 = \frac{\tilde{\beta}_1 - \hat{\beta}_1^{(z)}}{\hat{\beta}_z} = \frac{0.1366 - 0.1667}{-0.141} = 0.213$$

Ten more places go together with 2.1 more minutes of commuting. Municipalities well endowed with childcare are urban, with longer commutes, and commuting reduces hours ($\hat{\beta}_z < 0$). The bias in column (2) is $\hat{\beta}_z \hat{\delta}_1 = -0.0301$. It is negative, so column (2) underestimates the effect of places by about 18%.

Marks: 0.75 calculation of $\hat{\delta}_1$ + 0.25 interpretation + 0.5 direction and size of the bias = 1.5

D3. $\rho = (\beta_1 + \beta_6)/\beta_1$, estimated by $0.0906/0.1743 = 0.520$. Jacobian: $\partial\rho/\partial\beta_1 = -\beta_6/\beta_1^2 = 2.755$, $\partial\rho/\partial\beta_6 = 1/\beta_1 = 5.737$. With $\widehat{\mathrm{Var}}(\hat{\beta}_1) = 1.8769 \times 10^{-4}$, $\widehat{\mathrm{Var}}(\hat{\beta}_6) = 4.2025 \times 10^{-4}$ and the covariance in the note:

$$\widehat{\mathrm{Var}}(\hat{\rho}) = 0.00142 + 0.01383 - 0.00595 = 0.00931 \qquad \text{s.e.} = 0.096$$

The 95% confidence interval is $[0.331,\ 0.709]$. The value 0.5 lies inside the interval ($t = 0.21$), so claim (ii) is compatible with the data. The value 1 lies far outside ($t = -4.98$), so the effect is significantly smaller for mothers of two or more children.

Marks: 0.5 $\hat{\rho}$ + 0.5 Jacobian + 0.5 variance with the covariance + 0.5 conclusions about 0.5 *and* 1 = 2

D4. Expected elements: *major revision*. (1) Correct the income-pooling test. With the covariance it is rejected (C2, C3, C4), so column (4) cannot be the preferred specification and claims (iii) and (iv) must be withdrawn. (2) Add commuting time or municipal characteristics, since the effect of places is underestimated by about 18%. In addition, use standard errors valid under heteroskedasticity. Rejection would be excessive, because the main result (i) survives and becomes stronger.

Marks: 0.25 decision + 0.5 two priority corrections, correctly justified + 0.25 consistency with the previous parts = 1

Typical mistakes.

- In D1, dividing by $\sum_i x_{1i}^2$ instead of $\sum_i \hat{r}_i^2$.
- In D2, reversing the identity and concluding to a positive bias.
- In D3, forgetting the covariance (the standard error would be overstated).