Solutions : Économetrics 1

Midterm – 2025

Exercise 1 (7 points)

Indicate the single correct answer. Below, we always assume we have an i.i.d. sample $(X_i', Y_i)_{i=1,\dots,n}$ with $X_i \in \mathbb{R}^k$.

1. **Having an R^2 very close to 1 (e.g., 0.99) indicates :**
   (a) that the regressors are probably highly correlated with each other ;
   (b) **that the regressors explain Y very well in the sample considered ;**
   (c) that the regression coefficients are significant at usual levels (1% and 5%) ;
   (d) that there is probably a heteroskedasticity problem.

   Answer : (b)
   Justification : By definition, R^2 measures the share of variance of Y explained by the model. A value close to 1 means that fitted values $\hat{Y}$ are very close to observed values Y.
   — (a) is false.
   — (c) is false : global fit and individual significance are distinct.
   — (d) is false.

2. **We add to the initial sample a new observation (X_{n+1}', Y_{n+1}) satisfying $Y_{n+1} = X_{n+1}'\hat{\beta}_n$. Let $\hat{\beta}_{n+1}$ be the new OLS estimator. Then :**
   (a) $\hat{\beta}_{n+1}$ is not defined ;
   (b) $\hat{\beta}_{n+1} = \hat{\beta}_n$;
   (c) $\hat{\beta}_{n+1} \neq \hat{\beta}_n$ but the asymptotic variance estimate remains unchanged ;
   (d) $\hat{\beta}_{n+1} \neq \hat{\beta}_n$ and the asymptotic variance estimate changes.

   Answer : (b)
   Justification : The added observation has zero residual under $\hat{\beta}_n$, so the minimizer stays the same.

3. **We obtain $\hat{\beta}_j = 3.2$ with standard error 1.6. Using the normal table :**
   (a) We reject $H_0 : \beta_{0j} \leq 0$ against $H_1 : \beta_{0j} > 0$ at 1% ;
   (b) We reject $H_0 : \beta_{0j} \geq 0$ against $H_1 : \beta_{0j} < 0$ at 10% ;
   (c) The 95% confidence interval of β_{0j} includes 0 ;
   (d) **The p-value of the test $H_0 : \beta_{0j} \leq 0$ vs $H_1 : \beta_{0j} > 0$ is less than 2.5%.**

   Answer : (d)
   Justification : $t = 3.2/1.6 = 2$.

— (a) false.

— (b) false.

— (c) false.

— (d) true : $P(Z > 2) < 2.5\%$.

4. **We aim to predict Y_{n+1} using a subset of variables A. If $A = \{1, \dots, k\}$ (all variables), then :**
   (a) **The R^2 of this regression will be at least as large as that of regressions using $A' \subsetneq A$;**
   (b) The R^2 will be equal to 1 ;
   (c) The prediction $\hat{Y}_{n+1}^{A}$ will be better than $\hat{Y}_{n+1}^{A'}$ for all $A' \subsetneq A$;
   (d) The prediction $\hat{Y}_{n+1}^{A}$ will be worse than $\hat{Y}_{n+1}^{A'}$ for some $A' \subseteq A$.

Answer : (a)
Justification : Adding regressors weakly increases in-sample R^2. Predictive quality cannot be guaranteed.

5. **Suppose $X = (1, D)'$. Let $\hat{\beta}_D$ be the coefficient of D. Let $\tilde{D} = cD$ ($c > 1$). Then :**
   (a) $\hat{\beta}_{\tilde{D}} = \hat{\beta}_D$;
   (b) $\hat{\beta}_{\tilde{D}} = \hat{\beta}_D / c$**;**
   (c) $\hat{\beta}_{\tilde{D}} = \hat{\beta}_D + c$;
   (d) $\hat{\beta}_{\tilde{D}} = c\hat{\beta}_D$.

Answer : (b)
Justification : Predictions must match : $\hat{\beta}_D D = \hat{\beta}_{\tilde{D}}(cD)$.

6. **Consider the Ridge program** $\min_\beta \frac{1}{n} \sum (Y_i - X_i'\beta)^2 + \lambda \|\beta\|_2^2$**. This problem :**
   (a) has no solution if $n < k$;
   (b) has multiple solutions if $n > k$;
   (c) has multiple solutions if $\sum X_i X_i'$ is not invertible ;
   (d) **always has a unique solution.**

Answer : (d)
Justification : For $\lambda > 0$, $X'X + \lambda I$ is always invertible.

7. **Let $\hat{\beta}_D$ be the coefficient of D in the regression on (1, D, G) and $\hat{\beta}_D^S$ that in the regression on (1, D). Then :**
   (a) $\widehat{Cov}(Y, G) = 0$ implies $\hat{\beta}_D^S = \hat{\beta}_D$;
   (b) $\hat{\beta}_G > 0$ implies $\hat{\beta}_D > \hat{\beta}_D^S$;
   (c) $\widehat{Cov}(D, G) = 0$ **implies** $\hat{\beta}_D^S = \hat{\beta}_D$**;**
   (d) None of the above.

Answer : (c)
Justification : Omitted variable bias : $\hat{\beta}_D^S \approx \hat{\beta}_D + \hat{\beta}_G \frac{Cov(D,G)}{Var(D)}$.

Exercise 2 (9 points)

Study of wage (lwage) as a function of physical difficulties and education level.

1. **Interpretation of the coefficient of diffphysical and significance (Regression 1 : $\hat{\beta} = -0.1855$, $t = -13.72$).**

   The coefficient is -0.1855. We therefore predict a wage that is about **18.6% lower** for people who have difficulty walking (compared to those who do not).
   Significance : The test statistic $|t| = 13.72$ is far above the usual critical values of the normal distribution (2.57 for 1%). The coefficient is therefore significant at the 1%, 5%, and 10% levels.

2. **Relationship between diffphysical and nb_school after adding nb_school (Regression 2 : diffphysical coefficient falls to -0.1376).**

   The coefficient of `diffphysical` decreases in absolute value (from -0.1855 to -0.1376) when `nb_school` is introduced.
   From the omitted-variable formula :

   $$\hat{\beta}_D^{Short} \approx \hat{\beta}_D^{Long} + \hat{\beta}_{nb\_school} \times \frac{Cov(D, nb\_school)}{Var(D)}$$

   Here, $\hat{\beta}_{nb\_school} = 0.0924 > 0$. We observe that the short coefficient (-0.18) is more negative than the long one (-0.13). This implies that the bias term $\beta_{nb\_school} \times \lambda$ is negative. Since $\beta_{nb\_school} > 0$, this implies that the correlation between `diffphysical` and `nb_school` is **negative**.
   Interpretation : Individuals who have difficulty walking have, on average, fewer years of schooling than others.

3. **Prediction with an interaction term (Regression 3).**

   Analysis via the marginal effect (Comparison with the reference group) :

   The question asks whether, in this model, we predict a higher wage for people with walking difficulties compared to those without.
   The interaction term means that the effect of `diffphysical` is no longer constant but depends on education level. The **marginal effect** is :

   $$\frac{\partial \widehat{Y}}{\partial \text{diffphysical}} = \hat{\beta}_{diff} + \hat{\beta}_{inter} \times \text{nb\_school}$$

   Numerically :

   $$0.1767 - 0.0220 \times \text{nb\_school}$$

   Since nb_school ≥ 10, the most favorable (largest) marginal effect is :

   $$0.1767 - 0.0220(10) = -0.0433$$

   Conclusion : The marginal effect is always negative for all observed education levels (at best -4.33%). **No**, the model never predicts a higher wage for people with difficulties ; it always predicts a lower wage relative to the reference group (those without difficulties).

   Methodological remark (Intra-model vs. cross-model comparison) :
   One must not confuse marginal effects within a given model with changes in predicted wage across different model specifications.

- Comparing Regression 2 (no interaction) with Regression 3 (with interaction), one may observe that for low education levels (nb_school < 14.3), Regression 3 predicts a smaller penalty (e.g., around -4.3% instead of -13.7%).
- However, this type of cross-model comparison is not the correct interpretation in econometrics. When asked whether a variable predicts a "higher" value, the convention is to reason *ceteris paribus* relative to the **reference group** ($D = 0$) within the current specification.

Prediction for a person with difficulties and 12 years of schooling :

$$\hat{y} = \underbrace{1.9389}_{\hat{\alpha}} + \underbrace{0.1767}_{\hat{\beta}_{diff}} + \underbrace{0.0927(12)}_{\hat{\beta}_{school}\times 12} + \underbrace{(-0.0220)(12)}_{\hat{\beta}_{inter}\times 12}$$

This corresponds to a total effect of :

$$0.1767 - 0.264 = -0.0873,$$

that is, a penalty of about **8.7%**.

4. **Separate regression on the subgroup of individuals with difficulties.**

Yes, it is possible to determine these values. The interaction model is mathematically equivalent to running two separate regressions.
Demonstration (Parameter identification) : Consider the prediction equation of the full model :

$$\widehat{\text{lwage}} = \hat{\alpha} + \hat{\beta}_{diff} D + \hat{\beta}_{school} S + \hat{\beta}_{inter}(D \times S)$$

For individuals with difficulties ($D = 1$) :

$$\widehat{\text{lwage}}_{|D=1} = (\hat{\alpha} + \hat{\beta}_{diff}) + (\hat{\beta}_{school} + \hat{\beta}_{inter})S$$

We can identify the coefficients that would result from a separate regression of `lwage` on S for this subgroup :

- **Intercept :**
$$\hat{\alpha}_{sep} = 1.9389 + 0.1767 = \mathbf{2.1156}$$
- **Slope :**
$$\hat{\beta}_{sep} = 0.0927 - 0.0220 = \mathbf{0.0707}$$

Interpretation : For individuals with difficulties, the OLS regression line would have an intercept of 2.12 and a slope of 7.1%.

5. **Test of equality of coefficients (diffhear vs diffvision).**

We want to test the null hypothesis $H_0 : \beta_{hear} = \beta_{vision}$, which is a linear restriction $R\beta = 0$.
Method 1 : Rewriting the model and using a Student test
Define $\theta = \beta_{hear} - \beta_{vision}$. Substitute $\beta_{hear} = \theta + \beta_{vision}$ in the model :

$$Y = \cdots + \theta X_{hear} + \beta_{vision}(X_{hear} + X_{vision}) + \ldots$$

Testing H_0 is equivalent to testing if the coefficient of X_{hear} in this reparametrized model is zero.
The test statistic is :

$$t = \frac{\hat{\theta}}{\widehat{se}(\hat{\theta})} = \frac{\hat{\beta}_{hear} - \hat{\beta}_{vision}}{\widehat{se}(\hat{\beta}_{hear} - \hat{\beta}_{vision})}$$

Using the inequality provided :

$$\widehat{se}(\hat{\beta}_{hear} - \hat{\beta}_{vision}) \leq 0.0136 + 0.0176 = 0.0312$$

Hence :

$$|t| \geq \frac{|0.0255 - (-0.0910)|}{0.0312} = \frac{0.1165}{0.0312} \approx 3.73$$

Since $3.73 > 1.96$, we reject H_0 at the 5% level.

Method 2 : Wald / Fisher test (matrix formulation)
Let R select 1 for `diffhear` and -1 for `diffvision`. The Wald statistic is :

$$F = \frac{1}{r}(R\hat{\beta})' \left[R\hat{V}R'\right]^{-1} (R\hat{\beta})$$

With $r = 1$, this becomes :

$$F = \frac{(\hat{\beta}_{hear} - \hat{\beta}_{vision})^2}{\widehat{Var}(\hat{\beta}_{hear} - \hat{\beta}_{vision})} = t^2$$

Using the previous bound :

$$F \geq (3.73)^2 \approx 13.9$$

This far exceeds the 5% critical value for $F(1, \infty)$ (about 3.84), so we **reject the null hypothesis**.

6. **Including the variable difficulty (maximum of the 3 others).**

The variable `difficulty` equals 1 if at least one of the three variables (physical, hear, vision) equals 1.

— **Impossible case :** If the difficulties were mutually exclusive (no one has more than one problem), then `difficulty = diffphysical + diffhear + diffvision`, which would be perfect collinearity.

— **General case :** In reality, individuals may have multiple impairments. `difficulty` is therefore **not** a perfect linear combination of the others. **It can therefore be included** (no perfect collinearity).

Exercise 3 (5 points)

Log–linear model $Y = \exp(c + \beta_D D + \beta_G G + \epsilon)$.

1. **Estimator of β_D.**

The linearized model is

$$\ln(Y) = c + \beta_D D + \beta_G G + \epsilon.$$

By the Frisch–Waugh theorem, the estimator is

$$\hat{\beta}_D = \frac{\widehat{Cov}(r_D, \tilde{Y})}{\widehat{Var}(r_D)},$$

where r_D is the residual from regressing D on G (and a constant), and $\tilde{Y} = \ln(Y)$.

2. **Comment on the asymptotic variance and correlation.**

The statement is **FALSE**. Increasing the correlation between regressors leads to an *increase* (not a decrease) in the variance of the estimator.
Justification : From the course, the asymptotic variance of $\hat{\beta}_D$ in the multiple regression model is

$$V_a(\hat{\beta}_D) = \frac{V(\epsilon)}{V(D)\,(1 - R^2_\infty)},$$

where R^2_∞ is the limit of the R^2 obtained when regressing D on G.
In this simple setting,

$$R^2_{DG} = \text{Corr}(D, G)^2.$$

If $|\text{Corr}(D, G)|$ increases, then :

— $(1 - R^2_{DG})$ decreases toward 0,

— which makes $V_a(\hat{\beta}_D)$ **increase**, potentially diverging.

This is the classical **variance inflation due to multicollinearity**.

3. **Convergence of the estimator of the prediction $E[Y|X = x]$.**

The econometrician proposes

$$\hat{m}(x) = \exp(x'\hat{\beta}).$$

This estimator is **not consistent** for $E[Y|X = x]$.
Proof :

(a) *Limit of the proposed estimator :* Since $\hat{\beta} \xrightarrow{P} \beta_0$, the continuous mapping theorem gives :

$$\exp(x'\hat{\beta}) \xrightarrow{P} \exp(x'\beta_0).$$

(b) *True conditional mean :* The model is $Y = \exp(X'\beta_0 + \epsilon)$. Because ϵ is independent of X :

$$E[Y|X = x] = \exp(x'\beta_0)\, E[\exp(\epsilon)].$$

(c) *Use of Jensen's inequality :* Since the exponential function is strictly convex :

$$E[\exp(\epsilon)] > \exp(E[\epsilon]) = 1,$$

because $E(\epsilon) = 0$ and $V(\epsilon) > 0$.

(d) *Conclusion :* The estimator converges to $\exp(x'\beta_0)$, while the true value is

$$\exp(x'\beta_0) \times E[\exp(\epsilon)].$$

Since $E[\exp(\epsilon)] > 1$, the proposed estimator **systematically underestimates** $E[Y|X = x]$.

Consistent estimator : A consistent estimator must correct for $E[\exp(\epsilon)]$. Using the empirical mean of the exponentiated residuals :

$$\widehat{E}[Y|X = x] = \exp(x'\hat{\beta}) \times \left(\frac{1}{n} \sum_{i=1}^{n} \exp(\hat{\epsilon}_i) \right).$$