Econométrie 1

Corrigé de l'examen final, janvier 2025.

2 heures, sans document ni calculatrice.

Exercice 1 (5 points)

1 point par bonne réponse, -0,25 point par mauvaise réponse, 0 point si pas de réponse.

1. Soit $(Y_i, D_i)_{i=1,\dots,n}$ un échantillon et supposons $P(D=0)=P(D=1)=P(D=2)=1/3$. Soit $(\widehat{\alpha}, \widehat{\beta}_D)$ l'estimateur des moindres carrés ordinaires (MCO) de la régression de Y sur D, avec $\widehat{\beta}_D$ le coefficient relatif à D. Soit $\widehat{Y} = \widehat{\alpha} + \widehat{\beta}_D D$ et $\widehat{\varepsilon} = Y - \widehat{Y}$. Alors :
   - (a) Le coefficient de pente β_D de la régression théorique de Y sur D vérifie $\beta_D = E(Y|D=1) - E(Y|D=0)$.
   - (b) Si les $(D_i)_{i=1,\dots,n}$ sont non tous égaux, $\widehat{\text{Cov}}(\widehat{Y}, \widehat{\varepsilon}) = 0$.
   - (c) Si les $(D_i)_{i=1,\dots,n}$ sont non tous égaux, on peut avoir $\sum_{i=1}^n \widehat{\varepsilon}_i \neq \sum_{i=1}^n \widehat{\varepsilon}_i D_i$.
   - (d) L'estimateur $\widehat{\beta}_D$ ne converge pas nécessairement en probabilité car la condition d'inversibilité de $E[(1,D)(1,D)']$ peut ne pas être vérifiée sous les hypothèses énoncées.
2. Soit $(Y_i, X_i')_{i=1,\dots,n}$ l'échantillon observé, avec $X \in \mathbb{R}^k$ et $E(XX')$ inversible. Soit $\beta_0 \in \mathbb{R}^k$ les coefficients de la régression théorique de Y sur X, avec $\beta_0 \neq (0,\dots,0)' \in \mathbb{R}^k$. Soit $\varepsilon = Y - X'\beta_0$ le résidu de la régression théorique de Y sur X, et $E(\varepsilon^2 XX') \neq E(\varepsilon^2)E(XX')$. On note $\widehat{\beta}$ l'estimateur des MCO de la régression de Y sur X. Alors :
   - (a) $\sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}\left(0, E[\varepsilon^2]E[XX']^{-1}\right)$.
   - (b) $\|\sqrt{n}(\widehat{\beta} - \beta_0)\|^2 \xrightarrow{d} \chi^2_k$.
   - (c) $\|\sqrt{n}\widehat{\beta}\| \xrightarrow{P} \infty$.
   - (d) $\widehat{\beta}$ ne peut pas être calculé si $\widehat{V}(Y) = 0$.
3. Soit $(Y_i, X_i')_{i=1,\dots,n}$ l'échantillon observé et considérons le problème de minimisation
   $$\min_{\beta \in \mathbb{R}^k} \sum_{i=1}^n |Y_i - X_i'\beta|^2 + \lambda \|\beta\|^2$$
   où $X \in \mathbb{R}^k$, $\lambda > 0$, $\|\beta\|$ est la norme Euclidienne de β et $k \geq n$. Alors :
   - (a) Si $k > n$, la condition d'inversibilité sur $E(XX')$ ne tient pas, donc la solution n'est pas unique.
   - (b) La solution peut être ou non unique en fonction de la valeur de $\lambda > 0$.
   - (c) La solution est toujours unique et est « sparse », c'est-à-dire que la solution $\widehat{\beta}$ aura plusieurs de ses composantes égales à zéro.
   - (d) La solution est toujours unique mais n'est pas « sparse ».
4. Soit $Y_i(d)$ le résultat potentiel de l'individu i sous le statut de traitement $d \in \{0,1\}$, D_i le traitement de l'individu i et $Y_i = Y_i(D_i)$. On suppose $P(D=1) > 0$ et on note β_D le coefficient de la régression théorique de Y sur D. On observe un échantillon $(Y_i, D_i)_{i=1,\dots,n}$. Alors :

(a) $\text{Cov}(D, Y(0)) = 0$ implique $\beta_D = E[Y(1) - Y(0)]$.

(b) Pour tout $i \in \{1, \dots, n\}$, la probabilité d'observer $Y_i(1)$ est nulle.

(c) Pour estimer de manière convergente $E[Y(1) - Y(0)|D = 1]$, on doit supposer que les effets du traitement sont constants : $Y_i(1) - Y_i(0) = \delta$ pour un δ non aléatoire.

(d) $D \perp\!\!\!\perp (Y(0), Y(1))$ implique $\beta_D = E[Y(1) - Y(0)|D = 1]$.

5. Soit $Y_i(d)$ le résultat potentiel de l'individu i sous le statut de traitement $d \in \{0, 1\}$; $Z_i = 1$ si l'individu i est affecté au groupe de traitement et $Z_i = 0$ sinon ; soit $D_i(z) = 1$ si l'individu i prend le traitement sous l'affectation $z \in \{0, 1\}$ et $D_i(z) = 0$ sinon. On note $D_i = D_i(Z_i)$ et $Y_i = Y_i(D_i)$ le traitement et le résultat observés pour l'individu i. On suppose $V(Z) > 0$, $D(1) \geq D(0)$, $E(D|Z = 1) > E(D|Z = 0) > 0$ et $Z \perp\!\!\!\perp (Y(1), Y(0), D(1), D(0))$. Enfin, on observe $(Y_i, D_i, Z_i)_{i=1,\dots,n}$. Alors :

(a) $\widehat{\beta}_D \xrightarrow{P} E(Y|Z = 1) - E(Y|Z = 0)$, où $\widehat{\beta}_D$ est le coefficient de pente de la régression de Y sur D.

(b) Si l'on note D^* la prédiction linéaire théorique de D par Z, on a
$$\frac{E(Y|Z = 1) - E(Y|Z = 0)}{E(D|Z = 1) - E(D|Z = 0)} = \frac{\text{Cov}(Y, D^*)}{\text{V}(D^*)}.$$

(c) L'effet causal $E[Y(1) - Y(0)|D = 1]$ vérifie
$$E[Y(1) - Y(0)|D = 1] = \frac{E(Y|Z = 1) - E(Y|Z = 0)}{E(D|Z = 1) - E(D|Z = 0)}.$$

(d) On ne peut identifier aucun effet causal si $\text{Cov}(D, Y(0)) \neq 0$.

Corrigé

1-b, 2-c, 3-d, 4-d, 5-b.

Exercice 2 (10 points)

On cherche à mesurer les rendements de l'éducation des hommes aux Etats-Unis. On dispose pour ce faire d'une base de données où l'on mesure le logarithme du salaire (`lwage`), le nombre d'années d'éducation compté à partir de l'équivalent du CP (`educ`), le nombre d'années d'expérience (`exper`) et son carré (`expersq`), le fait d'être noir (`black`), de vivre dans le sud des Etats-Unis (`south`), d'habiter une ville de plus de 50 000 habitants (`city`), ou d'y avoir habité 10 ans auparavant (`city10`) [1].

On présente les statistiques descriptives suivantes :

1. Les données utilisées ici sont celles du « National Longitudinal Survey of Young Men »(NLSYM) de 1966.

Variable	Moyenne	Ecart-type
`lwage`	6,26	0,44
`educ`	13,26	2,68
`exper`	8,86	4,14
`expersq`	95,58	84,62

Table 1 – Moyennes de quelques variables

1. *(2 points)* On considère la régression ci-dessous. A quoi correspondent les colonnes « Estimate », « Std. Error » et « Pr(>|t|) » ? Comment peut-on interpréter le chiffre 0.093171 (ligne `educ`) ? Que penser de la phrase « il y a probablement un biais de sélection car le R^2 de la régression est faible » ?

```
Call:
lm_robust(formula = lwage ~ educ + exper + expersq, data = card)

Standard error type:  HC2

Coefficients:
             Estimate Std. Error t value   Pr(>|t|)  CI Lower  CI Upper   DF
(Intercept)  4.468541  0.0703272  63.539  0.000e+00  4.330646  4.606435 3006
educ         0.093171  0.0036805  25.315 2.503e-128  0.085954  0.100387 3006
exper        0.089783  0.0070933  12.657  8.217e-36  0.075875  0.103691 3006
expersq     -0.002486  0.0003414  -7.281  4.210e-13 -0.003155 -0.001816 3006

Multiple R-squared:  0.1958 , Adjusted R-squared:  0.195
F-statistic: 231.4 on 3 and 3006 DF,  p-value: < 2.2e-16
```

2. *(2 points)* On cherche à mesurer l'effet marginal moyen de l'expérience. Donner la formule théorique ainsi qu'un estimateur de ce paramètre. Peut-on calculer cet estimateur ici ? Si oui, remplacer les valeurs théoriques par les valeurs numériques (sans faire le calcul final), sinon indiquer les informations manquantes pour ce faire. Mêmes questions pour l'effet marginal moyen de l'expérience pour les individus ayant 12 année ou moins d'éducation.

3. *(1 points)* On considère maintenant la régression ci-dessous. Commenter la significativité statistique de `city` et `city10`. Commenter également l'évolution du coefficient d'`educ`. Pouvait-on s'attendre à une telle évolution ?

```
Call:
lm_robust(formula = lwage ~ educ + exper + expersq + south +
    black + city + city10, data = card)

Standard error type:  HC2

Coefficients:
             Estimate Std. Error t value  Pr(>|t|)  CI Lower  CI Upper   DF
(Intercept)  4.731360  0.0701260  67.469 0.000e+00  4.593860  4.868860 3002
educ         0.073850  0.0036446  20.263 1.135e-85  0.066704  0.080996 3002
exper        0.083516  0.0067402  12.391 2.007e-34  0.070300  0.096732 3002
expersq     -0.002246  0.0003185  -7.053 2.172e-12 -0.002871 -0.001622 3002
south       -0.123210  0.0153980  -8.002 1.736e-15 -0.153402 -0.093018 3002
black       -0.187532  0.0175203 -10.704 2.894e-26 -0.221885 -0.153179 3002
city         0.144433  0.0191052   7.560 5.329e-14  0.106972  0.181893 3002
city10       0.025528  0.0181646   1.405 1.600e-01 -0.010089  0.061144 3002

Multiple R-squared:  0.2909 , Adjusted R-squared:  0.2893
F-statistic:   187 on 7 and 3002 DF,  p-value: < 2.2e-16
```

4. *(1,5 points)* On cherche maintenant à utiliser une variable instrumentale pour `educ`. Expliquer pourquoi et donner le signe du biais qu'on peut anticiper sur l'estimateur du coefficient d'`educ` dans la question précente. On dispose des variables `IQ` (mesure de compétences cognitives), `nearc` (indicatrice d'habiter à proximité d'une université), `married` (indicatrice d'être marié), `whitecol` (indicatrice d'être un col blanc). Quelle variable semble la plus appropriée pour servir d'instrument ? Justifier sa réponse.
5. *(1,5 points)* On note `z` l'instrument retenu et on considère la régression ci-dessous. Pourquoi est-ce important d'étudier cette régression ? Indiquer quelle(s) hypothèse(s) nécessaire(s) à la validité de l'approche instrumentale on peut tester à partir de cette régression, et le(s) résultat(s) de ce(s) test(s) ici.

```
Call:
lm_robust(formula = educ ~ z + exper + expersq + south + black +
    city + city10, data = card)

Standard error type:  HC2

Coefficients:
              Estimate Std. Error  t value  Pr(>|t|)  CI Lower  CI Upper   DF
(Intercept) 16.6553648   0.147054 113.2602 0.000e+00 16.367028 16.943702 3002
z            0.3211497   0.084520   3.7997 1.477e-04  0.155427  0.486872 3002
exper       -0.4100488   0.032089 -12.7783 1.897e-36 -0.472968 -0.347129 3002
expersq      0.0007195   0.001711   0.4205 6.742e-01 -0.002636  0.004075 3002
south       -0.2897870   0.078555  -3.6890 2.292e-04 -0.443814 -0.135760 3002
black       -1.0008330   0.088216 -11.3452 3.059e-29 -1.173803 -0.827863 3002
city         0.3679973   0.111622   3.2968 9.893e-04  0.149134  0.586861 3002
city10       0.0617206   0.109529   0.5635 5.731e-01 -0.153039  0.276480 3002

Multiple R-squared:  0.4745 , Adjusted R-squared:  0.4733
F-statistic: 521.2 on 7 and 3002 DF,  p-value: < 2.2e-16
```

6. *(2 points)* Commenter la sortie ci-dessous et indiquer quel estimateur elle permet d'obtenir. Commenter l'évolution du coefficient d'`educ` par rapport à la question 3. Expliquer pourquoi cette évolution peut sembler surprenante et proposer une explication.

```
Call:
ivreg(formula = lwage ~ exper + expersq + south + black + city +
    city10 | educ | z, data = card)

Residuals:
     Min       1Q   Median       3Q      Max
-1.76404 -0.22980  0.02407  0.24842  1.40655

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.9768470  0.8988598   4.424 1.00e-05 ***
educ         0.1187744  0.0534800   2.221   0.0264 *
exper        0.1019455  0.0229290   4.446 9.06e-06 ***
expersq     -0.0022778  0.0003285  -6.933 5.02e-12 ***
south       -0.1083668  0.0235244  -4.607 4.26e-06 ***
black       -0.1429038  0.0560350  -2.550   0.0108 *
city         0.1266673  0.0294666   4.299 1.77e-05 ***
city10       0.0173992  0.0217012   0.802   0.4228
```

Corrigé

1. « Estimate » : valeurs de l'estimateur des MCO. « Std. Error » : écarts-types des estimateurs des MCO. « Pr(>|t|) » : p-valeur des tests des hypothèses $\beta_{0j} = 0$ où $\beta_0 = (\beta_{01}, ..., \beta_{0k})'$ est le vecteur de la régression théorique (1 point). Le nombre 0.093171 signifie que toutes choses égales par ailleurs (l'expérience, en l'occurrence), on prédit qu'un an d'étude supplémentaire est associé à un salaire plus élevé de 9,3% (0.5 point). Enfin, la phrase en question n'a pas de sens car le R^2 mesure la qualité de prédiction de la régression et ne donne pas d'indication sur l'aspect causal de la régression (0,5 pt).

2. Notons δ_0 l'effet marginal moyen de l'expérience. Il s'agit de l'effet sur la variation de moyenne des salaires qu'on prédit si tous les individus connaissaient un changement infinitésimal d'expérience. Si l'on note $f_\beta(\texttt{exper})$ la fonction de l'expérience introduite dans la régression, on a $\delta_0 = E[f'_\beta(\texttt{exper})]$. Ici $f_\beta(\texttt{exper}) = \texttt{exper}\beta_{\texttt{exper}} + \texttt{exper}^2\beta_{\texttt{expersq}}$. Donc

   $$\delta_0 = \beta_{\texttt{exper}} + 2E[\texttt{exper}]\beta_{\texttt{expersq}} \quad \text{(0.5 point).}$$

   On peut estimer δ_0 par

   $$\widehat{\delta} = \widehat{\beta}_{\texttt{exper}} + 2\overline{\texttt{exper}}\widehat{\beta}_{\texttt{expersq}},$$

   où $\overline{\texttt{exper}}$ est la moyenne sur l'échantillon de l'expérience (0.5 point). A partir des sorties et des statistiques descriptives, on a ici :

   $$\widehat{\delta} = 0,09 + 2 \times 8,86 \times (-0,0025) \quad \text{(0.25 point).}$$

   En raisonnant comme précédemment, l'effet marginal moyen de l'expérience pour les individus ayant 12 années ou moins d'éducation (δ_1 disons) est

   $$\delta_1 = \beta_{\texttt{exper}} + 2E[\texttt{exper}|\texttt{educ} \leq 12]\beta_{\texttt{expersq}} \quad \text{(0,25 point).}$$

   Un estimateur est

   $$\widetilde{\delta} = \widehat{\beta}_{\texttt{exper}} + 2\overline{\texttt{exper}}_{\texttt{educ}\leq 12}\widehat{F\beta}_{\texttt{expersq}},$$

où $\overline{\texttt{exper}}_{\texttt{educ}\leq 12}$ est la moyenne sur l'échantillon de l'expérience pour les individus ayant 12 années ou moins d'éducation (0.25 point). On ne peut pas calculer $\widetilde{\delta}$ ici car $\overline{\texttt{exper}}_{\texttt{educ}\leq 12}$ n'est pas disponible (0.25 point).

3. Le coefficient de `city` est significatif aux seuils habituels (y compris 0,1%). Le coefficient de `city10` n'est pas significatif aux seuils habituels (y compris 10%) (0,5 point). L'ajout de variables de contrôle a conduit à une diminution du coefficient de la variable `educ`. Cette évolution semble logique : dans la régression précédente, on peut penser que l'effet causal de l'éducation était surestimé, du fait de variables omises corrélées négativement sur le salaire et l'éducation, comme la variable `black`. Le lien entre coefficient des régressions courtes et longues montre que la diminution du coefficient d'`educ` était alors attendue (0,5 point).

4. On peut penser qu'il subsiste un biais de variables omises, lié par exemple aux capacités cognitives qui joueraient à la fois sur l'éducation des individus et sur leur salaire (0,25 point). Si tel est le cas, on anticipe une suresimation du coefficient des MCO par rapport au véritable effet causal de l'éducation (0,25 point). On peut y remédier en adoptant une stratégie instrumentale. Il s'agit ici d'utiliser une variable corrélée à l'éducation (condition de pertinence) mais n'ayant pas d'effet direct sur le salaire (relation d'exclusion). En ce sens la variable `IQ` n'est a priori pas valide. Il ne semble pas évident que la variable `married` soit corrélée à l'éducation. Comme la variable `IQ`, la variable `whitecol` a a priori un impact direct sur le salaire donc elle n'est a priori pas valide non plus. La variable `nearc`, en revanche, semble valide : habiter à proximité d'une université pourrait favoriser le fait d'y aller (car les coûts seraient réduits) et n'a a priori pas d'effet direct sur le salaire (1 point : 0,25 point pour le bon choix de l'instrument, 0,75 point pour la justification).

5. Il s'agit ici de la régression de première étape des doubles moindres carrés. Il est important d'étudier cette régression car celle-ci permet de tester la condition de pertinence (0,5 point). Sans celle-ci, l'estimateur des doubles moindres carrés n'est pas convergent (0,25 point). Cette condition de pertinence correspond au fait que le coefficient de `z` dans la régression doit être différent de zéro (0,5 point). Ici il semble que oui, puisqu'on rejette la nullité du coefficient à tous les seuils habituels (y compris 0,1%) (0,25 point).

6. La sortie permet d'obtenir l'estimateur des doubles moindres carrés, où `educ` est instrumenté par `z` (0,5 point). On voit que les écarts-types de cette régression sont beaucoup plus grands que dans la régression de la question 3. Ceci pouvait être attendu car la variance de l'estimateur des doubles moindres carrés est en général beaucoup plus élevée que celle de l'estimateur des moindres carrés ordinaires (0,5 point). On constate aussi que le coefficient d'`educ` a augmenté, ce qui est contraire à ce à quoi on pouvait s'attendre, cf. la réponse à la question 4 (0,25 point). Une explication possible serait des erreurs de mesure sur le niveau d'éducation, puisqu'on a alors un biais d'atténuation. Une autre explication serait que les rendements de l'éducation sont hétérogènes, et qu'on mesure avec l'estimateur des doubles moindres carrés un effet local (LATE) correspondant aux individus dont l'éducation est affectée par l'instrument (i.e., ils ne poursuivent leur éducation que parce qu'ils vivent à proximité d'une université). Ces individus auraient des rendements plus élevés que les autres, ce qui expliquerait l'augmentation (0,75 point pour une explication valide).

Exercice 3

Soit $Y_i(d)$ le résultat potentiel de l'individu i sous le statut de traitement d. On suppose le modèle suivant pour le résultat potentiel :

$$Y_i(d) = c_0 + \Delta_i d + \lambda_0 Z_i + \eta_i \text{ avec } E(\eta_i) = 0 \text{ et } E(\Delta_i|Z_i, D_i) = \delta_0\,, \tag{0.1}$$

où c_0 et λ_0 sont des coefficients fixes. On suppose également que

$$D_i(z) = \alpha_0 + \gamma_0 z + U_i + \zeta_0 \eta_i \text{ où } E(U_i) = 0 \text{ et } U_i \perp\!\!\!\perp \eta_i\,. \tag{0.2}$$

On suppose enfin que

$$Z_i \perp\!\!\!\perp (U_i, \eta_i) \text{ et } \mathrm{V}(Z_i) > 0\,.$$

Soit $D_i := D_i(Z_i)$ et $Y_i := Y_i(D_i)$. On observe un échantillon i.i.d. $(Y_i, D_i, Z_i)_{i=1,\dots,n}$.

1. *(1 point)* A partir de l'équation (0.1), obtenir une équation causale liant Y à $(1, D, Z)$ (chacune multipliée par un coefficient fixe) et un terme d'erreur de moyenne nulle.
2. *(1 point → 1,5 point)* Montrer que si $\zeta_0 \neq 0$, l'estimateur des MCO de δ_0 correspondant à la régression ci-dessus n'est pas convergent.
3. *(1 point)* Supposons que $\zeta_0 \neq 0$. Sous quelles conditions sur λ_0 et γ_0 peut-on estimer δ_0 de manière convergente ? Sous de telles conditions, proposer un estimateur $\widehat{\delta}$ de δ_0 et, en utilisant les résultats vus dans le cours, justifier qu'il est convergent.
4. On veut tester l'hypothèse nulle d'absence d'effet moyen du traitement contre l'alternative d'un effet moyen positif, c'est-à-dire

$$H_0 : \delta_0 = 0 \text{ contre } H_1 : \delta_0 > 0\,.$$

   (a) *(0.75 points → 1,5 point)* On rappelle que

$$\sqrt{n}\left(\widehat{\delta} - \delta_0\right) \xrightarrow{d} \mathcal{N}(0, V_a)$$

   pour un certain V_a. Donner l'expression de V_a et proposer un estimateur convergent de celui-ci.

   (b) *(0.5 points)* Proposer une statistique de test pour tester H_0 contre H_1.

   (c) *(0.75 points)* Préciser la valeur critique et la règle de décision pour tester H_0 contre H_1.

Corrigé

1. On obtient à partir de l'équation (0.1)

$$Y_i = c_0 + D_i\delta_0 + \lambda Z_i + r_i,$$

   avec $r_i := (\Delta_i - \delta_0)D_i + \eta_i$ (0,25 point). De plus,

$$E[r_i|D_i] = (E[\Delta_i|D_i] - \delta_0)D_i + E[\eta_i|D_i] = 0,$$

   car $E[\Delta_i|D_i, Z_i] = \delta_0$ et $E[\eta_i|D_i, Z_i] = 0$ impliquent $E[\Delta_i|D_i] = \delta_0$ et $E[\eta_i|D_i] = 0$ (0,75 point).

2. Il suffit de constater que dans la relation causale ci-dessus, le terme d'erreur r_i est corrélé à D_i (0,5 point). Pour le voir, notons déjà que $E[D_i r_i] = E[D_i(\Delta_i - \delta_0)D_i] + E[D_i \eta_i]$. De plus,

$$E[D_i(\Delta_i - \delta_0)D_i] = E[D_i^2(E(\Delta_i|D_i) - \delta_0)] = 0.$$

   Par ailleurs, $D_i = \alpha_0 + \gamma_0 Z_i + U_i + \zeta_0 \eta_i$. Donc,

$$E[D_i\eta_i] = \alpha_0 E[\eta_i] + \gamma_0 E[Z_i\eta_i] + E[U_i\eta_i] + \zeta_0 V(\eta_i) = \zeta_0 V(\eta_i) \neq 0,$$

   où l'on a utilisé $E[\eta_i] = 0$, $Z_i \perp\!\!\!\perp \eta_i$ et $U_i \perp\!\!\!\perp \eta_i$ (1 point). Donc l'estimateur des MCO ne converge pas vers δ_0 lorsque $\zeta_0 \neq 0$.

3. Notons que $E[Z_i r_i] = 0$. L'idée alors est que sous certaines conditions, Z pourrait servir d'instrument à D (0,25 point). Il faut pour ce faire que les conditions de pertinence et d'exogénéité soit vérifiée. Cette dernière l'est si $\lambda_0 = 0$ (Z_i n'a alors pas d'effet direct sur Y_i) (0,25 point). La condition de pertinence est que $\text{Cov}(D_i, Z_i) \neq 0$, soit encore $\gamma_0 \neq 0$ (0,25 point). L'estimateur des doubles moindres carrés (2MC), où D est instrumenté par Z, vérifiant ces deux conditions, il est alors convergent, d'après le cours (0,25 point).

4. (a) L'estimateur $\widehat{\delta}$ est asymptotiquement équivalent à l'estimateur des MCO de la régression infaisable de Y sur $D^* := \alpha_0 + \gamma_0 Z$ (cf. la preuve du théorème 6 du Chapitre 5) :

$$\sqrt{n}\left(\widehat{\delta} - \delta_0\right) = \frac{1}{V(D^*)}\frac{1}{\sqrt{n}}\sum_{i=1}^n D_i^* r_i + o_P(1).$$

   Par conséquent, $V_a = E[(r_i D^*)^2]/V(D^*)^2$ (1 point). On peut l'estimer de manière convergente par

$$\widehat{V}_a = \frac{\frac{1}{n}\sum_{i=1}^n \widehat{r}_i^2 \widehat{D}_i^2}{\frac{1}{n}\sum_{i=1}^n \left(\widehat{D}_i - \overline{\widehat{D}}\right)^2},$$

   avec $\widehat{D}_i := \widehat{\alpha} + \widehat{\gamma} Z_i$ et $\widehat{r}_i := Y_i - \widehat{c} - D_i\widehat{\delta}$ (où $\widehat{c}$ est l'estimateur des 2MC de la constante c_0) (0,5 point).

   (b) On peut alors utiliser la statistique de test $T = \widehat{\delta}/\widehat{V}_a^{1/2}$ (0,5 point).

   (c) Sous l'hypothèse nulle $T \xrightarrow{d} \mathcal{N}(0,1)$ tandis que sous l'hypothèse alternative, $T \xrightarrow{P} \infty$. Donc, pour obtenir un test convergent, on considère, pour obtenir un test de niveau asymptotique α, la région critique $\{T > q_{1-\alpha}\}$ avec $q_{1-\alpha}$ le quantile d'ordre $1-\alpha$ d'une loi normale standard (0,75 point).