Université d'Orléans - Master Econométrie et Statistique Appliquée
Econométrie des Variables Qualitatives

Christophe HURLIN

Correction Examen Mai 2011. C. Hurlin

Exercice 1 (15 points) : Modèle Tobit

Question 1 *(1 point)* : On sait que l'on a :

$$\Pr(y_i = 0) = \Pr(y_i^* < 0) = 1 - \Phi\left(\frac{\alpha + x_i\beta}{\sigma_\varepsilon}\right) \quad \text{1 point} \tag{1}$$

Question 2 *(2 points)* : Compte tenu de l'échantillon, les contributions à la vraisemblance sont les suivantes de chaque type d'individus sont les suivantes :

Nombre d'Observations	y_i	x_i	Contribution
9 observations	$y_i = 0$	$x_i = 1$	$1 - \Phi\left(\frac{\alpha+\beta}{\sigma_\varepsilon}\right)$
3 observations	$y_i = 0$	$x_i = 0$	$1 - \Phi\left(\frac{\alpha}{\sigma_\varepsilon}\right)$
1 observation	$y_i = \sqrt{4}$	$x_i = 0$	$\frac{1}{\sigma_\varepsilon}\phi\left(\frac{\sqrt{4}}{\sigma_\varepsilon}\right)$
1 observation	$y_i = \sqrt{2}$	$x_i = 0$	$\frac{1}{\sigma_\varepsilon}\phi\left(\frac{\sqrt{2}}{\sigma_\varepsilon}\right)$
1 observation	$y_i = 1$	$x_i = 1$	$\frac{1}{\sigma_\varepsilon}\phi\left(\frac{-1-\beta}{\sigma_\varepsilon}\right)$

Dès lors, la vraisemblance s'écrit

$$\begin{aligned}
\log L\left(y;\alpha,\beta,\sigma_{\varepsilon}^{2}\right) &= \sum_{i:\,y_i=0}\log\left[1-\Phi\left(\frac{x_i\beta}{\sigma_{\varepsilon}}\right)\right]+\sum_{i:\,y_i>0}\log\left[\left(\frac{1}{\sigma_{\varepsilon}}\right)\phi\left(\frac{y_i-x_i\beta}{\sigma_{\varepsilon}}\right)\right] \\
&= 9\log\left[1-\Phi\left(\frac{\alpha+\beta}{\sigma_{\varepsilon}}\right)\right]+3\log\left[1-\Phi\left(\frac{\alpha}{\sigma_{\varepsilon}}\right)\right] \\
&\quad +\log\left[\frac{1}{\sigma_{\varepsilon}}\phi\left(\frac{\alpha+\sqrt{4}}{\sigma_{\varepsilon}}\right)\right]+\log\left[\frac{1}{\sigma_{\varepsilon}}\phi\left(\frac{\alpha+\sqrt{2}}{\sigma_{\varepsilon}}\right)\right]+\log\left[\frac{1}{\sigma_{\varepsilon}}\phi\left(\frac{\alpha+1-\beta}{\sigma_{\varepsilon}}\right)\right] \\
&= 9\log\left[1-\Phi\left(\frac{\alpha+\beta}{\sigma_{\varepsilon}}\right)\right]+3\log\left[1-\Phi\left(\frac{\alpha}{\sigma_{\varepsilon}}\right)\right]-\frac{3}{2}\log\left(\sigma_{\varepsilon}^{2}\right) \\
&\quad +\log\left[\phi\left(\frac{\alpha+\sqrt{4}}{\sigma_{\varepsilon}}\right)\right]+\log\left[\phi\left(\frac{\alpha+\sqrt{2}}{\sigma_{\varepsilon}}\right)\right]+\log\left[\phi\left(\frac{\alpha+1-\beta}{\sigma_{\varepsilon}}\right)\right]
\end{aligned}$$

En simplifiant, il vient :

$$\begin{aligned}
\log L\left(y;\alpha,\beta,\sigma_{\varepsilon}^{2}\right) &= 9\log\left[1-\Phi\left(\frac{\alpha+\beta}{\sigma_{\varepsilon}}\right)\right]+3\log\left[1-\Phi\left(\frac{\alpha}{\sigma_{\varepsilon}}\right)\right]-\frac{3}{2}\log\left(\sigma_{\varepsilon}^{2}\right) \\
&\quad -\frac{3}{2}\log\left(2\pi\right)-\frac{1}{2\sigma_{\varepsilon}^{2}}\left[\left(\alpha+\sqrt{4}\right)^{2}+\left(\alpha+\sqrt{2}\right)^{2}+\left(1+\alpha-\beta\right)^{2}\right] \qquad (2)
\end{aligned}$$

ou encore

$$\begin{aligned}
\log L\left(y;\alpha,\beta,\sigma_{\varepsilon}^{2}\right) &= 9\log\left[1-\Phi\left(\frac{\alpha+\beta}{\sigma_{\varepsilon}}\right)\right]+3\log\left[1-\Phi\left(\frac{\alpha}{\sigma_{\varepsilon}}\right)\right]-\frac{3}{2}\log\left(\sigma_{\varepsilon}^{2}\right) \\
&\quad -\frac{3}{2}\log\left(2\pi\right)-\frac{1}{2\sigma_{\varepsilon}^{2}}\left(7+2\beta+\beta^{2}\right) \qquad (3)
\end{aligned}$$

Question 3 *(2 points)* : On pose z_i une variable dichotomique telle que :

$$z_i=\begin{cases}1 & \text{si } y_i>0 \\ 0 & \text{sinon}\end{cases} \qquad \forall i=1,..N \qquad (4)$$

La vraisemblance du modèle probit associé à z_i s'écrit :

$$\begin{aligned}
\log L\left(z;\frac{\alpha}{\sigma_{\varepsilon}},\frac{\beta}{\sigma_{\varepsilon}}\right) &= \sum_{i=1}^{N} z_i\log\left[\Phi\left(\frac{\alpha+x_i\beta}{\sigma_{\varepsilon}}\right)\right]+\left(1-z_i\right)\log\left[1-\Phi\left(\frac{\alpha+x_i\beta}{\sigma_{\varepsilon}}\right)\right] \\
&= 2\log\left[\Phi\left(\frac{\alpha}{\sigma_{\varepsilon}}\right)\right]+\log\left[\Phi\left(\frac{\alpha+\beta}{\sigma_{\varepsilon}}\right)\right] \\
&\quad +9\log\left[1-\Phi\left(\frac{\alpha+\beta}{\sigma_{\varepsilon}}\right)\right]+3\log\left[1-\Phi\left(\frac{\alpha}{\sigma_{\varepsilon}}\right)\right]
\end{aligned}$$

Question 4 *(2 points)* : On cherche les estimateurs du MV $\left(\widehat{\alpha}/\widehat{\sigma}_{\varepsilon},\widehat{\beta}/\widehat{\sigma}_{\varepsilon}\right)$ tels que :

$$\left(\widehat{\alpha}/\widehat{\sigma}_{\varepsilon},\widehat{\beta}/\widehat{\sigma}_{\varepsilon}\right)=\underset{\left\{\frac{\alpha}{\sigma_{\varepsilon}},\frac{\beta}{\sigma_{\varepsilon}}\right\}}{\arg\max}\left[\log L\left(z;\frac{\alpha}{\sigma_{\varepsilon}},\frac{\beta}{\sigma_{\varepsilon}}\right)\right] \qquad (5)$$

Considérons le changement de variable suivant :

$$z_1=\Phi\left(\frac{\alpha+\beta}{\sigma_{\varepsilon}}\right) \qquad z_2=\Phi\left(\frac{\alpha}{\sigma_{\varepsilon}}\right) \qquad (6)$$

La log-vraisemblance s'écrit alors :

$$\begin{aligned}\log L(z;z_1,z_2) &= 2\times\log(z_2)+\log(z_1)\\ &\quad +9\times\log(1-z_1)+3\times\log(1-z_2)\end{aligned}\tag{7}$$

On cherche alors le couple $(\widehat{z}_1,\widehat{z}_2)$ tel que :

$$(\widehat{z}_1,\widehat{z}_2)=\underset{\{z_1,z_2\}}{\arg\max}\log L(z;z_1,z_2)\tag{8}$$

Les conditions nécessaires s'écrivent :

$$\left.\frac{\partial L(z;z_1,z_2)}{\partial z_1}\right|_{\widehat{z}_1}=\frac{1}{\widehat{z}_1}-\frac{9}{1-\widehat{z}_1}=0\tag{9}$$

$$\left.\frac{\partial L(z;z_1,z_2)}{\partial z_2}\right|_{\widehat{z}_2}=\frac{2}{\widehat{z}_2}-\frac{3}{1-\widehat{z}_2}=0\tag{10}$$

D'où l'on tire que :

$$\widehat{z}_1=\frac{1}{10}\quad \widehat{z}_2=\frac{2}{5}\tag{11}$$

On en déduit alors les estimateurs de MV $\widehat{\alpha}$ et $\widehat{\beta}$ puisque :

$$\frac{\widehat{\alpha}}{\widehat{\sigma}_\varepsilon}=\Phi^{-1}(\widehat{z}_2)=\Phi^{-1}\left(\frac{2}{5}\right)=-0.2533\tag{12}$$

$$\frac{\widehat{\beta}}{\widehat{\sigma}_\varepsilon}=\Phi^{-1}(\widehat{z}_1)-\Phi^{-1}(\widehat{z}_2)=\Phi^{-1}\left(\frac{1}{10}\right)-\Phi^{-1}\left(\frac{2}{5}\right)=-1.0282\tag{13}$$

Question 5 *(2 points)* : Déterminons la valeur du ratio de Mills pour les trois individus non censurés. Par définition :

$$\lambda\left(\frac{\widehat{\alpha}+x_i\widehat{\beta}}{\widehat{\sigma}_\varepsilon}\right)=\phi\left(\frac{\widehat{\alpha}+x_i\widehat{\beta}}{\widehat{\sigma}_\varepsilon}\right)\left[\Phi\left(\frac{\widehat{\alpha}+x_i\widehat{\beta}}{\widehat{\sigma}_\varepsilon}\right)\right]^{-1}\tag{14}$$

Il vient donc pour les deux individus non censurés pour lesquels $x_i=0$, on a :

$$\begin{aligned}\lambda\left(\frac{\widehat{\alpha}}{\widehat{\sigma}_\varepsilon}\right) &= \phi\left(\frac{\widehat{\alpha}}{\widehat{\sigma}_\varepsilon}\right)\left[\Phi\left(\frac{\widehat{\alpha}}{\widehat{\sigma}_\varepsilon}\right)\right]^{-1}\\ &= \phi(-0.2533)\left[\Phi(-0.2533)\right]^{-1}\\ &= 0.9658\end{aligned}\tag{15}$$

Pour l'individu non censuré pour lesquels $x_i=1$, on a :

$$\begin{aligned}\lambda\left(\frac{\widehat{\alpha}}{\widehat{\sigma}_\varepsilon}\right) &= \phi\left(\frac{\widehat{\alpha}}{\widehat{\sigma}_\varepsilon}\right)\left[\Phi\left(\frac{\widehat{\alpha}}{\widehat{\sigma}_\varepsilon}\right)\right]^{-1}\\ &= \phi(-0.2533-1.0282)\left[\Phi(-0.2533-1.0282)\right]^{-1}\\ &= 1.7549\end{aligned}\tag{16}$$

Question 6 *(2 points)* : La régression de la deuxième étape d'Heckman ne porte que sur les trois individus non censurés et s'écrait sous la forme :

$$y_i=x_i\beta+\sigma_\varepsilon\,\lambda\left(\frac{\widehat{\alpha}+x_i\widehat{\beta}}{\widehat{\sigma}_\varepsilon}\right)+v_i\tag{17}$$

On a donc ici :

$$\begin{pmatrix} \sqrt{4} \\ \sqrt{2} \\ -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0.9658 \\ 1 & 0 & 0.9658 \\ 1 & 1 & 1.7549 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \sigma_\varepsilon \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \tag{18}$$

L'estimateur des MCO des paramètres α, β et σ est donc défini par :

$$\begin{aligned} \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \\ \widehat{\sigma}_\varepsilon \end{pmatrix} &= \left[\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0.9658 & 0.9658 & 1.7549 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0.9658 \\ 1 & 0 & 0.9658 \\ 1 & 1 & 1.7549 \end{pmatrix} \right]^{-1} \\ &\quad \times \left[\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0.9658 & 0.9658 & 1.7549 \end{pmatrix} \times \begin{pmatrix} \sqrt{4} \\ \sqrt{2} \\ 1 \end{pmatrix} \right] \end{aligned} \tag{19}$$

On obtient $\widehat{\alpha} = 0.9696$, $\widehat{\beta} = -3.1544$ et $\widehat{\sigma}_\varepsilon = 0.5353$.

Question 7 *(2 points)* : On admet que les estimateurs OLS des paramètres de la regression linéaire suivante :

$$y_i = \alpha + x_i \beta + \mu_i \tag{20}$$

obtenus à partir des données portant **sur l'ensemble** des 15 individus de l'échantillon, sont $\widehat{\alpha} = 0.3104$ et $\widehat{\beta} = -0.5604$. On sait que ces estimateurs OLS sont biaisés et que sous les hypothèses de Goldberger, on a :

$$\widehat{\beta}_{LS} \underset{N \to \infty}{\overset{p}{\longrightarrow}} \beta \times \Phi\left(\frac{\alpha}{\sigma_y}\right) \tag{21}$$

où α correspond à la constante de l'équation $y_i^* = \alpha + x_i\beta + \varepsilon_i$ et $\sigma_y^2 = \sigma_\varepsilon^2 + \beta'\Omega\beta$, où Ω désigne la matrice de variance covariance des variables explicatives x_i. Un estimateur non biaisé de β (idem pour α) est donc défini par la quantité $\widehat{\beta}_{LS}^c = (N/N_1) \times \widehat{\beta}_{LS}$, où N_1 est le nombre d'observations pour lesquelles $y_i^* > 0$, est un estimateur convergent de β.

$$\widehat{\beta}_{LS}^c = \frac{N}{N_1} \widehat{\beta}_{LS} \underset{N \to \infty}{\overset{p}{\longrightarrow}} \beta \tag{22}$$

Ici $N_1 = 3$ et $N = 15$, on a donc :

$$\widehat{\beta}_{LS}^c = -0.5604 \times \frac{15}{3} = -2.801 \tag{23}$$

$$\widehat{\alpha}_{LS}^c = 0.3104 \times \frac{15}{3} = 1.5519 \tag{24}$$

Ces valeurs sont sensiblement éloignées de celles obtenus par la méthode d'Heckman, puisque les hypothèses de Goldberger ne sont pas satisfaites dans notre cas, x_i étant une variable dichotomique, qui ne peut être normalement distribuée (point bonus + 1).

Question 8 *(2 points)* : Dans notre cas, on a :

$$\frac{\partial E\left(y_i / x_i\right)}{\partial x_i} = \Phi\left(\frac{\alpha + x_i\beta}{\sigma_\varepsilon}\right)\beta \qquad \forall i = 1, .., N \tag{25}$$

Pour les individus tels que $x_i = 0$, on a :

$$\begin{aligned} \frac{\partial E\left(y_i / x_i\right)}{\partial x_i} &= \Phi\left(\frac{\alpha}{\sigma_\varepsilon}\right)\beta \\ &= \Phi\left(\frac{0.9696}{0.5353}\right) \times (-3.1544) \\ &= -3.0439 \end{aligned} \tag{26}$$

Pour les individus tels que $x_i = 1$, on a :

$$\begin{aligned}
\frac{\partial E\left(y_i/\,x_i\right)}{\partial x_i} &= \Phi\left(\frac{\alpha+\beta}{\sigma_\varepsilon}\right)\beta \\
&= \Phi\left(\frac{0.9696-3.1544}{0.5353}\right)\times(-3.1544) \\
&= -0.000007 \qquad (27)
\end{aligned}$$

Exercice 2 (6 points) : Modele Logit Indépendant

Question 1 *(2 points)* : cf. cours.

Question 2 *(2 points)* : Pour chaque variable, y compris le terme constant nommé intercept, ne figurent que trois paramètres. Celui relatif à la catégorie de référence a été omis puisqu'on sait qu'il vaut 0. La catégorie de référence est ici les ouvriers (cs = 4). Les trois paramètres de chaque variable sont associés aux trois catégories cs, hors la catégorie de référence, classées dans l'ordre croissant des modalités de la variable cs. Par exemple, le paramètre age_1 (0.0873) mesure l'effet de âge sur l'appartenance à la catégorie des cadres (plutôt qu'à celle des ouvriers). De façon générale, on a :

$$\Pr\left(y_i = j\right) == \frac{\exp\left(x_i\beta_j\right)}{1+\sum\limits_{k=1}^{3}\exp\left(x_i\beta_k\right)} \qquad (28)$$

où le vecteur β_4 est normalisé à zéro : $\beta_4 = 0$. Puisque le modèle logit indépendant satisfait l'hypothèse IIA, on a pour deux modalités j et k distinctes:

$$\frac{\Pr\left(y_i = j\right)}{\Pr\left(y_i = k\right)} = \frac{\exp\left(x_i\beta_j\right)}{\exp\left(x_i\beta_k\right)} = \exp\left[x_i\left(\beta_j-\beta_k\right)\right] \qquad (29)$$

En particulier pour les modalités cadre ($j = 1$) et ouvrier ($k = 4$), on a :

$$\frac{\Pr\left(y_i = 1\right)}{\Pr\left(y_i = 4\right)} = \frac{\exp\left(x_i\beta_1\right)}{\exp\left(x_i\beta_4\right)} = \exp\left[x_i\left(\beta_1-\beta_4\right)\right] = \exp\left(x_i\beta_1\right) \qquad (30)$$

Dès lors pour une femme ($sexe = 1$), de nationalité française ($etr = 0$), de 30 ans ($age = 30$), ayant fini ses études à 22 ans ($afinet = 22$), on a :

$$\begin{aligned}
\frac{\Pr\left(y_i = 1\right)}{\Pr\left(y_i = 4\right)} &= \exp\left(-15.3794+0.0873*age-0.2628*fem+0.6157*afnet-1.5111*etr\right) \\
&= \exp\left(-15.3794+0.0873*30-0.2628*1+0.6157*22-1.5111*0\right) \\
&= 1.6857 \qquad (31)
\end{aligned}$$

Cette femme a 1.68 fois plus de chances d'être cadre qu'ouvrière.

Question 3 *(2 points)* : L'effet marginal de l'âge sur la probabilité que l'individu i choisisse la $j^{ème}$ modalité, $\forall j = 1, .., 4$, est :

$$\frac{\partial p_{i,j}}{\partial age_i} = p_{i,j}\left[\beta_j^{age}-\sum_{z=1}^{4}p_{i,z}\beta_z^{age}\right] \qquad (32)$$

On sait que pour cet individu :

$$\Pr(y_i = 1) = 0.1380 \tag{33}$$

$$\Pr(y_i = 2) = 0.2297 \tag{34}$$

$$\Pr(y_i = 3) = 0.2209 \tag{35}$$

$$\Pr(y_i = 4) = 0.4114 \tag{36}$$

Dès lors, il vient :

$$\begin{aligned}
\frac{\Pr(y_i = 1)}{\partial age_i} &= \Pr(y_i = 1) \times \left[0.0873 - \sum_{z=1}^{4} \Pr(y_i = j) \times \beta_z^{age}\right] \\
&= 0.1380 \times [0.0873 - 0.1380 \times 0.0873 - 0.2297 \times 0.0595 \\
&\quad -0.2209 \times 0.0375 - 0.4114 \times 0] \\
&= 0.0074
\end{aligned} \tag{37}$$

$$\begin{aligned}
\frac{\Pr(y_i = 4)}{\partial age_i} &= \Pr(y_i = 4) \times \left[0 - \sum_{z=1}^{4} \Pr(y_i = j) \times \beta_z^{age}\right] \\
&= 0.4114 \times [0 - 0.1380 \times 0.0873 - 0.2297 \times 0.0595 \\
&\quad -0.2209 \times 0.0375 - 0.4114 \times 0] \\
&= -0.0140
\end{aligned} \tag{38}$$

Pour cet individu, lorsque l'âge augmente, la probabilité d'être cadre augmente et celle d'être ouvrier diminue.