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

Christophe HURLIN

Correction Examen Mai 2009. C. Hurlin

Exercice 1 (13 points) : Modèle Probit[1]

Question 1 *(1 point)* : Sous ces hypothèse, la probabilité d'apparition de l'évenement $y_i = 1$ s'écrit :

$$\begin{aligned} \Pr(y_i = 1) &= \Pr(y_i^* \geq 0) \\ &= \Pr(\varepsilon_i \geq -\alpha - \beta x_i) \\ &= 1 - \Pr(\varepsilon_i < -\alpha - \beta x_i) \end{aligned} \tag{1}$$

Sachant que ε_i $N.i.d.(0,1)$, et que la loi normale est symétrique, on obtient finalement :

$$\Pr(y_i = 1) = \Pr(\varepsilon_i < \alpha + \beta x_i) \tag{2}$$

ou encore

$$\Pr(y_i = 1) = \Phi(\alpha + \beta x_i) \quad \text{(1 point)} \tag{3}$$

où $\Phi(.)$ désigne la fonction de répartition de la loi normale centrée réduite.

Question 2 *(2 points)* : La log-vraisemblance de ce modèle probit s'écrit de la façon habituelle :

$$\log L(y; \alpha, \beta) = \sum_{i=1}^{N} y_i \log[\Phi(\alpha + x_i\beta)] + (1 - y_i) \log[1 - \Phi(\alpha + x_i\beta)] \tag{4}$$

Dans le cadre de cette application, on obtient alors :

$$\begin{aligned} \log L(y; \alpha, \beta) = \; & 16 \times \log[\Phi(\alpha + \beta)] + 26 \times \log[\Phi(\alpha)] \\ & + 32 \times \log[1 - \Phi(\alpha + \beta)] + 26 \times \log[1 - \Phi(\alpha)] \end{aligned} \tag{5}$$

Question 3 *(2 points)* : On cherche les estimateurs du MV $\left(\widehat{\alpha}, \widehat{\beta}\right)$ tels que :

$$\left(\widehat{\alpha}, \widehat{\beta}\right) = \underset{\{\alpha,\beta\}}{\arg\max} \left[\log L(y; \alpha, \beta)\right] \tag{6}$$

[1] D'après l'exemple donné dans le polycopié de cours de Jean Marc Robin "Econométrie des Variables Qualitatives", Université Paris 1.

Considérons le changement de variable suivant :

$$z_1 = \Phi(\alpha + \beta) \quad z_2 = \Phi(\alpha) \tag{7}$$

La log-vraisemblance s'écrit alors :

$$\begin{aligned} \log L(y; z_1, z_2) &= 16 \times \log(z_1) + 26 \times \log(z_2) \\ &\quad +32 \times \log(1 - z_1) + 26 \times \log(1 - z_2) \end{aligned} \tag{8}$$

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

$$(\widehat{z}_1, \widehat{z}_2) = \underset{\{z_1, z_2\}}{\arg\max} \left[\log L(y; z_1, z_2)\right] \tag{9}$$

Les conditions nécessaires s'écrivent :

$$\left.\frac{\partial L(y; z_1, z_2)}{\partial z_1}\right|_{\widehat{z}_1} = \frac{16}{\widehat{z}_1} - \frac{32}{1 - \widehat{z}_1} = 0 \tag{10}$$

$$\left.\frac{\partial L(y; z_1, z_2)}{\partial z_2}\right|_{\widehat{z}_2} = \frac{26}{\widehat{z}_2} - \frac{26}{1 - \widehat{z}_2} = 0 \tag{11}$$

D'où l'on tire que :

$$\widehat{z}_1 = \frac{1}{3} \quad \widehat{z}_2 = \frac{1}{2} \tag{12}$$

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

$$\widehat{\alpha} = \Phi^{-1}(\widehat{z}_2) = \Phi^{-1}(0.5) = 0 \tag{13}$$

$$\widehat{\beta} = \Phi^{-1}(\widehat{z}_1) - \Phi^{-1}(\widehat{z}_2) = \Phi^{-1}(1/3) - \Phi^{-1}(0.5) = -0.43 \tag{14}$$

Bonus : condition nécessaire (+ 1 point). La matrice jacobienne dans le système (z_1, z_2) est définie négative, puisque :

$$\begin{aligned} \left.\begin{pmatrix} \frac{\partial^2 L(y;z_1,z_2)}{\partial z_1^2} & \frac{\partial^2 L(y;z_1,z_2)}{\partial z_1 \partial z_2} \\ \frac{\partial^2 L(y;z_1,z_2)}{\partial z_1 \partial z_2} & \frac{\partial^2 L(y;z_1,z_2)}{\partial z_2^2} \end{pmatrix}\right|_{(\widehat{z}_1, \widehat{z}_2)} &= \begin{pmatrix} -\frac{16}{\widehat{z}_1^2} - \frac{32}{(1-\widehat{z}_1)^2} & 0 \\ 0 & -\frac{26}{\widehat{z}_2^2} - \frac{26}{(1-\widehat{z}_2)^2} \end{pmatrix} \\ &= \begin{pmatrix} -216 & 0 \\ 0 & -208 \end{pmatrix} \end{aligned} \tag{15}$$

Question 4 *(2 points)* : Déterminons l'effet marginal associé à la variable x. Sachant que la variable x_i est dichotomique, cet effet marginal s'écrit sous la forme suivante :

$$\varepsilon_{p_i|x_i} = \Pr(y_i = 1|\, x_i = 1) - \Pr(y_i = 1|\, x_i = 0) \tag{16}$$

ou encore

$$\varepsilon_{p_i|x_i} = \Phi(\alpha + \beta) - \Phi(\alpha) \tag{17}$$

On vérifie d'ores et déjà, qu'étant donné la nature dichotomique de la variable x_i et la spécification du modèle, l'effet marginal de la variable x_i est le même pour tous les individus $i = 1, .., N$ de l'échantillon. L'estimation de l'effet marginal vaut :

$$\begin{aligned} \widehat{\varepsilon}_{p_i|x_i} &= \Phi\left(\widehat{\alpha} + \widehat{\beta}\right) - \Phi(\widehat{\alpha}) \\ &= \widehat{z}_1 - \widehat{z}_2 \\ &= \frac{1}{3} - \frac{1}{2} \\ &= -\frac{1}{6} \end{aligned} \tag{18}$$

Question 5 *(2 points)* : On cherche à l'hypothèse nulle $H_0 : \beta = 0$ en utilisant un test de ratio de vraisemblance. On doit donc considérer la statistique suivante :-où $L(y; \widehat{\alpha}, 0)$ désigne la log-vraisemblance du modèle probit obtenue sous l'hypothèse nulle $H_0 : \beta = 0$. On sait que sous l'hypothèse alternative, on a :

$$\widehat{\alpha} = 0 \quad \widehat{\beta} = \Phi^{-1}(1/3) = -0.43 \tag{19}$$

Dès lors la valeur de la log-vraisemblance sous l'hypothèse alternative est :

$$\begin{aligned}
\log L\left(y; \widehat{\alpha}, \widehat{\beta}\right) &= \log L(y; \widehat{z}_1, \widehat{z}_2) \\
&= 16 \times \log(\widehat{z}_1) + 26 \times \log(\widehat{z}_2) \\
&\quad +32 \times \log(1 - \widehat{z}_1) + 26 \times \log(1 - \widehat{z}_2) \\
&= 16 \times \log(1/3) + 32 \times \log(2/3) + 2 \times 26 \times \log(1/2) &(20) \\
&= -66.59 &(21)
\end{aligned}$$

Sous l'hypothèse nulle la log-vraisemblance s'écrit de la façon suivante :

$$\begin{aligned}
\log L(y; \alpha, 0) &= 16 \times \log[\Phi(\alpha)] + 26 \times \log[\Phi(\alpha)] \\
&\quad +32 \times \log[1 - \Phi(\alpha)] + 26 \times \log[1 - \Phi(\alpha)] \\
&= 42 \times \log[\Phi(\alpha)] + 58 \times \log[1 - \Phi(\alpha)] &(22)
\end{aligned}$$

On pose $z = \Phi(\alpha)$, il vient :

$$\log L(y; z, 0) = 42 \times \log(z) + 58 \times \log(1 - z)$$

Dès lors la maximisation de la vraisemblance donne :

$$\left.\frac{\partial \log L(y; z, 0)}{\partial z}\right|_{\widehat{z}} = \frac{42}{\widehat{z}} - \frac{58}{1 - \widehat{z}} = 0 \tag{23}$$

D'om l'on tire que :

$$\widehat{z} = \frac{42}{100} = 0.42 \tag{24}$$

Par conséquent, la valeur de la log-vraisemblance sous l'hypothèse nulle est :

$$\begin{aligned}
\log L(y; \widehat{z}, 0) &= 42 \times \log(\widehat{z}) + 58 \times \log(1 - \widehat{z}) \\
&= 42 \times \log(0.42) + 58 \times \log(1 - 0.42) \\
&= -68.02 &(25)
\end{aligned}$$

On en déduit la valeur de la statistique LRT :

$$\begin{aligned}
LRT &= -2 \times \left[\log L(y; \widehat{\alpha}, 0) - \log L\left(y; \widehat{\alpha}, \widehat{\beta}\right)\right] \\
&= -2 \times (-68.02 - 66.59) \\
&= 2.86 &(26)
\end{aligned}$$

Sous H_0, cette statistique suit une loi du chi-deux à 1 degré de liberté. Pour un risque de première espèce, le seuil critique est égal à 3.84, donc on ne peut rejetter $H_0 : \beta = 0$.

Question 6 *(2 point)* : Par définition, on a :

$$R^2 \text{ de McFadden (1974) } = 1 - \frac{\log L\left(y, \widehat{\alpha}, \widehat{\beta}\right)}{\log L(y, \alpha, 0)} \tag{27}$$

Donc ici, on a :

$$R^2 \text{ de McFadden (1974)} = 1 - \frac{-66.59}{-68.02} = 0.021 \tag{28}$$

Question 7 *(2 points)* : Calculons les probabilités estimées de réalisation de l'événement $y_i = 1$ à partir de la formule suivante :

$$\widehat{\Pr}(y_i = 1) = \Phi\left(\widehat{\alpha} + \widehat{\beta} x_i\right) = \Phi(-0.43 x_i)$$

Nombre d'Observations	y_i	x_i	$\widehat{\Pr}(y_i = 1)$
16 observations	$y_i = 1$	$x_i = 1$	$\frac{1}{3}$
26 observations	$y_i = 1$	$x_i = 0$	$\frac{1}{2}$
32 observations	$y_i = 0$	$x_i = 1$	$\frac{1}{3}$
26 observations	$y_i = 0$	$x_i = 0$	$\frac{1}{2}$

ROC Curve cf. cours

Exercice 2 (8 points) : Modele Logit Conditionnel

Question 1 *(0.5 point)* : Modèle logit conditionnel. cf. cours

Question 2 *(0.5 point)* : En supposant qu'il n'existe pas de données manquantes, pour chaque individu et pour chaque variable explicative, sont reportées quatre valeurs correspondant aux quatre modalités (avion, train, bus et voiture). Sachant qu'il y 210 individus dans la base, on doit donc avoir pour chaque variable explicative un total de 840 observations.

Question 3 *(1 point).* Dans cette spécification on observe que $\beta_4 = 0$. Si $\beta_j > 0$, cela signifie que toutes choses égales par ailleurs, la probabilité de prendre la modalité de transport j est supérieure la probabilité de prendre la voiture.

Question 4 *(1 point) :* Dans le cadre du modèle logit conditionnel, on montre que pour l'individu n°2 :

$$x_{2,1}\widehat{\beta} = -2.540 \quad x_{2,2}\widehat{\beta} = -2.155 \quad x_{2,3}\widehat{\beta} = -3.771 \quad x_{2,4}\widehat{\beta} = -1.171 \tag{29}$$

avec

$$\begin{aligned} x_{2,1}\widehat{\beta} \quad = \quad & \widehat{\beta}_1 * \mathbb{I}_{(j=1)} + \widehat{\beta}_2 * \mathbb{I}_{(j=2)} + \widehat{\beta}_3 * \mathbb{I}_{(j=3)} + \widehat{\beta}_4 * tps\_trans_{2,j} \\ & + \widehat{\beta}_5 * tps\_trajet_{2,j} + \widehat{\beta}_6 * cout_{2,j}, \end{aligned} \tag{30}$$

Les probabilités estimées sont définies par :

$$\widehat{\Pr}(y_2 = j) = \frac{\exp\left(x_{2,j}\widehat{\beta}\right)}{\sum_{k=1}^{4} \exp\left(x_{2,k}\widehat{\beta}\right)} \tag{31}$$

On montre alors que :

$$\widehat{\Pr}(y_2 = 1) = 0.1495 \quad \widehat{\Pr}(y_2 = 2) = 0.2196 \quad \widehat{\Pr}(y_2 = 3) = 0.0436 \quad \widehat{\Pr}(y_2 = 4) = 0.5873 \tag{32}$$

Question 5 *(2 points)* : L'EM associé au cout de transport (noté par simplification $w_{i,j}$) est défini par :

$$\begin{aligned}\frac{\partial \Pr(y_i = j)}{\partial w_{i,j}} &= \frac{\partial}{\partial w_{i,j}}\left(\frac{\exp(x_{i,j}\beta)}{\sum_{k=1}^{4}\exp(x_{i,k}\beta)}\right) \\ &= \left[\beta_6 * \exp(x_{i,j}\beta) * \sum_{k=1}^{4}\exp(x_{i,k}\beta) - \beta_6 * \exp(x_{i,j}\beta) * \exp(x_{i,j}\beta)\right] \\ &\quad * \left[\sum_{k=1}^{4}\exp(x_{i,k}\beta)\right]^{-2} \qquad (33)\end{aligned}$$

On obtient alors :

$$\begin{aligned}\frac{\partial \Pr(y_i = j)}{\partial w_{i,j}} &= \beta_6 * \left\{\frac{\exp(x_{i,j}\beta)}{\left[\sum_{k=1}^{4}\exp(x_{i,k}\beta)\right]^{-2}}\right\} * [1 - \exp(x_{i,j}\beta)] \qquad (34) \\ &= \qquad (35)\end{aligned}$$

Pour l'individu 2, la modalité voiture ($j = 4$), l'effet maginal du coût de transport vaut :

$$\begin{aligned}\frac{\partial \Pr(y_i = 4)}{\partial w_{i,j}} &= -0.0139 * \left\{\frac{\exp(-1.171)}{\left[\sum_{k=1}^{4}\exp(x_{i,k}\beta)\right]^{-2}}\right\} * [1 - \exp(-1.171)] \qquad (36) \\ &= -0.0056 \qquad (37)\end{aligned}$$

On vérifie que dans ce cas, une hausse du coût du voyage diminue la probabilité de prendre la voiture.

Question 6 *(2 points)* : Pour un individu i, la log vraisemblance associée à un modèle logit conditionnel s'écrit alors :

$$\log L(y, \beta) = \sum_{j=1}^{4} y_{i,j} x_{i,j}\beta - \log\left[\sum_{k=1}^{4}\exp(x_{i,j}\beta)\right] \qquad (38)$$

Ici on a donc pour l'individu 2 :

$$\begin{aligned}\log L(y, \beta) &= \sum_{j=1}^{4} y_{2,j} x_{2,j}\beta - \log\left[\sum_{k=1}^{4}\exp(x_{2,j}\beta)\right] \qquad (39) \\ &= x_{2,4}\beta - \log\left[\sum_{k=1}^{4}\exp(x_{2,j}\beta)\right] \qquad (40)\end{aligned}$$

puisque l'individu 2 a choisit au final de prendre la voiture ($y_{2,4} = 1$, $y_{2,j} = 0$, pour $j \neq 4$). Dès lors, pour la valeur estimée des paramètres on obtient :

$$\begin{aligned}\log L\left(y, \widehat{\beta}\right) &= x_{2,4}\widehat{\beta} - \log\left[\sum_{k=1}^{4}\exp\left(x_{2,j}\widehat{\beta}\right)\right] \qquad (41) \\ &= -1.171 - \log\left[\sum_{k=1}^{4}\exp\left(x_{2,j}\widehat{\beta}\right)\right] \qquad (42) \\ &= -0.532 \qquad (43)\end{aligned}$$