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

Examen Mai 2011. C. Hurlin

Exercice 1 (15 points) : Modèle Tobit

On considère un modèle tobit tel que

$$y_i^* = \alpha + x_i\beta + \varepsilon_i \qquad \forall i = 1, ..N \tag{1}$$

$$y_i = \begin{cases} y_i^* & \text{si } y_i^* > 0 \\ 0 & \text{si } y_i^* \leq 0 \end{cases} \tag{2}$$

où y_i^* désigne une variable latente, x_i désigne une variable dichotomique prenant deux valeurs $\{0; 1\}$, α et β désignent deux réels. On suppose que les erreurs ε_i sont *i.i.d.* et suivent une loi $N\left(0, \sigma_\varepsilon^2\right)$. On dispose d'un échantillon de 15 observations des variables x_i et y_i (dont 12 sont censurées), réparties de la façon suivante :

Nombre d'Observations	y_i	x_i
9 observations	$y_i = 0$	$x_i = 1$
3 observations	$y_i = 0$	$x_i = 0$
1 observation	$y_i = \sqrt{4}$	$x_i = 0$
1 observation	$y_i = \sqrt{2}$	$x_i = 0$
1 observation	$y_i = 1$	$x_i = 1$

Question 1 *(1 point)* : Déterminez en fonction des paramètres α, β et σ_ε, la probabilité d'apparition de l'évenement de censure, $y_i = 0$.

Question 2 *(2 points)* : Montrez que la fonction de log-vraisemblance de ce modèle tobit peut s'écrire sous la forme :

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

où $\Phi(.)$ désigne la fonction de répartition de la loi $N(0,1)$.

Question 3 *(2 points)* : On se propose à présent d'estimer les paramètres α, β et σ_ε par la méthode d'Heckman en deux étapes. Soit z_i une variable dichotomique telle que :

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

Montrez que la vraisemblance du modèle probit associé à z_i, ne dépend que des ratios $\alpha/\sigma_\varepsilon$ et β/σ_ε et s'écrit :

$$\begin{aligned}\log L\left(z;\frac{\alpha}{\sigma_\varepsilon},\frac{\beta}{\sigma_\varepsilon}\right) &= 2\log\left[\Phi\left(\frac{\alpha}{\sigma_\varepsilon}\right)\right]+\log\left[\Phi\left(\frac{\alpha+\beta}{\sigma_\varepsilon}\right)\right]\\ &\quad +9\log\left[1-\Phi\left(\frac{\alpha+\beta}{\sigma_\varepsilon}\right)\right]+3\log\left[1-\Phi\left(\frac{\alpha}{\sigma_\varepsilon}\right)\right] \qquad (5)\end{aligned}$$

Question 4 *(2 points)* : Montrez que les réalisations des estimateurs du maximum de vraisemblance des paramètres normalisés $\alpha/\sigma_\varepsilon$ et β/σ_ε, obtenus à partir du modèle probit, valent :

$$\frac{\widehat{\alpha}}{\widehat{\sigma}_\varepsilon}=\Phi^{-1}\left(\frac{2}{5}\right) \qquad (6)$$

$$\frac{\widehat{\beta}}{\widehat{\sigma}_\varepsilon}=\Phi^{-1}\left(\frac{1}{10}\right)-\Phi^{-1}\left(\frac{2}{5}\right) \qquad (7)$$

Question 5 *(2 points)* : Déterminez la valeur des ratios de Mills, nécessaires pour la seconde étape de la démarche d'Heckman, pour les trois individus non censurés.

Question 6 *(2 points)* : Ecrivez sous forme vectorielle le modèle de régression de la deuxième étape d'Heckman en indiquant les valeurs de y_i, de x_i et du ratio de Mills obtenus à la question précédente. Montrez que les estimateurs d'Heckman des paramètres α, β et σ_ε vérifient :

$$\begin{aligned}\begin{pmatrix}\widehat{\alpha}\\ \widehat{\beta}\\ \widehat{\sigma}_\varepsilon\end{pmatrix} &= \left[\begin{pmatrix}1 & 1 & 1\\ 0 & 0 & 1\\ 0.9658 & 0.9658 & 1.7549\end{pmatrix}\times\begin{pmatrix}1 & 0 & 0.9658\\ 1 & 0 & 0.9658\\ 1 & 1 & 1.7549\end{pmatrix}\right]^{-1}\\ &\quad\times\left[\begin{pmatrix}1 & 1 & 1\\ 0 & 0 & 1\\ 0.9658 & 0.9658 & 1.7549\end{pmatrix}\times\begin{pmatrix}\sqrt{4}\\ \sqrt{2}\\ 1\end{pmatrix}\right] \qquad (8)\end{aligned}$$

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

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

$$y_i=\alpha+x_i\beta+\mu_i \qquad (9)$$

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 suppose (à tort, puisque x_i est une variable dichotomique) que les hypothèses de Goldberger (1981) sont valides. Proposez des estimateurs non biaisés des paramètres α et β, et comparez les à ceux obtenus à la question 6. Commentez.

Question 8 *(2 points)* : En assimilant la variable x_i à une variable continue, déterminez une approximation de l'effet marginal de cette variable sur la variable observée y_i pour tous les individus de l'échantillon, en distinguant ceux pour lesquels $x_i=0$ et ceux pour lesquels $x_i=1$. Commentez.

Exercice 2 (6 points) : Modele Logit Indépendant

On cherche à expliquer la catégorie professionnelle (CS) en fonction l'âge (variable age), de l'âge de fin d'etudes en années révolues (afinet), du sexe (variable binaire fem qui vaut 1 pour une femme et 0 pour un homme) et de la nationalité (variable etr, variable binaire qui vaut 1 si l'individu est de nationalité étrangère, 0 sinon). La CS est codée de la façon suivante: $cs = 1$ si le salarié est cadre, $cs = 2$ s'il exerce une profession intermédiaire, $cs = 3$ s'il est employé, $cs = 4$ s'il est ouvrier (catégorie de référence).

Question 1 *(2 points)* : Commentez précisèment tous les élements du tableau de résultats d'estimation (figures 1, 2 et 3)

Question 2 *(2 points)* : Calculez l'odds ratio (rapport de risques relatifs) associé au fait d'occuper un emploi de cade plutôt qu'un emploi d'ouvrier pour une femme de nationalité française, agée de 30 ans, ayant fini ses études à 22 ans.

Question 3 *(2 points)* : Calculez les effets marginaux de l'âge sur la probabilité d'être cadre ($cs = 1$) et ouvrier ($cs = 4$), pour un homme de nationalité française, agé de 30 ans ayant fini ses études à 19 ans. On admet que pour cet individu, les probabilités associées aux différentes CS sont les suivantes :

$$\Pr(y_i = 1) = 0.1380 \tag{10}$$

$$\Pr(y_i = 2) = 0.2297 \tag{11}$$

$$\Pr(y_i = 3) = 0.2209 \tag{12}$$

$$\Pr(y_i = 4) = 0.4114 \tag{13}$$

The QLIM Procedure

Multinomial Logit Estimates

Algorithm converged.

Model Fit Summary

Dependent Variable	csb
Number of Observations	2102
Log Likelihood	-2205
Maximum Absolute Gradient	5.99425E-6
Number of Iterations	6
Optimization Method	Newton-Raphson
AIC	4441
Schwarz Criterion	4526

Discrete Response Profile

Index	csb	Frequency	Percent
0	0	479	22.79
1	1	367	17.46
2	2	382	18.17
3	3	874	41.58

Figure 1: Source : Les modèles logit polytomiques non ordonnés : théorie et applications, Cédric AFSA ESSAFI, INSEE

Goodness-of-Fit Measures for Discrete Choice Models

Measure	Value	Formula
Likelihood Ratio (R)	1123.7	2 * (LogL - LogL0)
Upper Bound of R (U)	5534.6	- 2 * LogL0
Aldrich-Nelson	0.3484	R / (R+N)
Cragg-Uhler 1	0.4141	1 - exp(-R/N)
Cragg-Uhler 2	0.4462	(1-exp(-R/N)) / (1-exp(-U/N))
Estrella	0.4498	1 - (1-R/U)^(U/N)
Adjusted Estrella	0.4399	1 - ((LogL-K)/LogL0)^(-2/N*LogL0)
McFadden's LRI	0.2030	R / U
Veall-Zimmermann	0.4807	(R * (U+N)) / (U * (R+N))
McKelvey-Zavoina	0.6172	

N = # of observations, K = # of regressors

Figure 2: Source : Les modèles logit polytomiques non ordonnés : théorie et applications, Cédric AFSA ESSAFI, INSEE

Parameter Estimates

Parameter	DDL	Estimation	Standard Error	t Value	Approx Pr > |t|	Gradient
Intercept_1	1	-15.3794	0.8790	-17.50	<.0001	1.65E-7
age_1	1	0.0873	0.0112	7.80	<.0001	5.994E-6
afinet_1	1	0.6157	0.0315	19.52	<.0001	2.982E-6
fem_1	1	-0.2628	0.1737	-1.51	0.1303	7.393E-8
etr_1	1	-1.5110	0.3991	-3.79	0.0002	1.662E-7
Intercept_2	1	-10.5093	0.7887	-13.33	<.0001	-4.75E-8
age_2	1	0.0595	0.0103	5.78	<.0001	-1.73E-6
afinet_2	1	0.4285	0.0286	14.97	<.0001	-9.57E-7
fem_2	1	0.2758	0.1535	1.80	0.0723	-3.07E-8
etr_2	1	-1.1147	0.3283	-3.40	0.0007	-2.51E-8
Intercept_3	1	-4.5683	0.6318	-7.23	<.0001	-1.15E-7
age_3	1	0.0375	0.008625	4.35	<.0001	-4.25E-6
afinet_3	1	0.1485	0.0227	6.53	<.0001	-2.48E-6
fem_3	1	1.9408	0.1310	14.82	<.0001	-8.98E-8
etr_3	1	-0.3242	0.1998	-1.62	0.1045	-6.42E-8

Figure 3: Source : Les modèles logit polytomiques non ordonnés : théorie et applications, Cédric AFSA ESSAFI, INSEE