Université d'Orléans - Maitrise d'Econométrie
Econométrie des Variables Qualitatives

Examen Terminal Decembre 2003. C. Hurlin

Exercice 1 (12 points) : Modèle Tobit Simple Censuré

Le but de cet exercice est d'évaluer l'impact d'une action marketing sur le chiffre d'affaire d'une société de cosmétique. Cette société dispose pour cela de données d'enquête concernant la consommation annuelle du type de produit cosmétique concerné, notées c_i, récoltées auprès d'un échantillon de 10 000 individus. Pour chaque individu, indicé i, on dispose d'observations concernant:

- son revenu annuel moyen désigné par la variable r_i, exprimé en milliers d'euros
- le prix moyen des biens en euros relevé à partir des différents achats effectués par l'individu i, noté p_i.
- s'il a été ou non l'objet dans l'année d'une opération marketing. La variable dichotomique correspondante est codée $m_i = 1$ s'il y a une action marketing et $m_i = 0$ sinon.
- la catégorie socio-professionnelle (CSP) de l'individu représentée par une variable polytomique s_i codée respectivement 1 pour "chômeurs", 2 pour un statut "cadres et cadres supérieurs, 3 pour "retraités" et 0 pour le statut "ouvriers et employés" qui sera considéré comme référence.

On considère alors le modèle tobit simple censuré suivant :

$$c_i = \begin{cases} c_i^* & \text{si } c_i^* > 0 \\ 0 & \text{si } c_i^* \leq 0 \end{cases} \tag{1}$$

où c_i désigne une variable inobservable telle que :

$$c_i^* = \beta_0 + \beta_1 r_i + \beta_2 p_i + \beta_3 m_i + \beta_4 s_i + \varepsilon_i \tag{2}$$

où ε_i est $N.i.d.\left(0, \sigma_\varepsilon^2\right)$.

Partie I : Analyse de la spécification et modèle probit (7 points)

Question 1 *(2 points)* : (i) Précisez le signe attendu des paramètres théoriques β_j, $j = 0, 1, 2, 3$ en justifiant économiquement votre réponse. Que peut on conclure quant au signe de β_4 ? On définit des variables dichotomiques $csp_i^{(j)}$ associées à la CSP de l'individu i :

$$csp_i^{(j)} = \begin{cases} 1 & \text{si } s_i = j \\ 0 & \text{sinon} \end{cases} \quad j = 0, 1, 2, 3, \quad i = 1, .., N$$

(ii) Proposez une écriture du modèle (2) en substituant la variable s_i par autant de variables dichotomiques $csp_i^{(j)}$ qu'il est nécessaire et commentez les signes des coefficients γ_j associés à ces variables.

Question 2 *(1.5 points)* : On construit une variable dichotomique z_i valant 1 si la consommation observée de l'individu i est strictement positive et 0 dans le cas contraire. Dérivez la probabilité que l'agent i consomme effectivement un produit cosmétique de cette gamme en fonction des vecteurs de paramètres $\beta = (\beta_1\, \beta_2\, \beta_3)'$ et $\gamma = (\gamma_1\, \gamma_2\, \gamma_3)'$, d'une constante θ et de la variance des résidus σ_ε^2.

Question 3 *(1.5 points)* : Les paramètres du modèle probit associé à la probabilité de consommer ont été estimés à partir des 10 000 individus de l'échantillon par maximum de vraisemblance. Le résultat de ces estimations est reproduit sur la figure (1). Commentez ces résultats.

Figure 1: Résultats d'Estimation du Modèle Probit

Dependent Variable: Z
Method: ML - Binary Probit
Date: 12/09/03 Time: 10:39
Sample: 1 10000
Included observations: 10000
Convergence achieved after 12 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
R	3.199342	0.188923	16.93465	0.0000
P	-2.346068	0.144414	-16.24548	0.0000
M	1.507599	0.175970	8.567360	0.0000
CSP1	-1.777251	0.202092	-8.794280	0.0000
CSP2	5.624149	0.406264	13.84358	0.0000
CSP3	3.231677	0.281552	11.47808	0.0000
C	0.392022	0.411851	0.951855	0.3412

Mean dependent var	0.437300	S.D. dependent var	0.496078
S.E. of regression	0.082739	Akaike info criterion	0.046283
Sum squared resid	68.40928	Schwarz criterion	0.051330
Log likelihood	-224.4133	Hannan-Quinn criter.	0.047991
Restr. log likelihood	-6852.639	Avg. log likelihood	-0.022441
LR statistic (6 df)	13256.45	McFadden R-squared	0.967252
Probability(LR stat)	0.000000		

Obs with Dep=0	5627	Total obs	10000
Obs with Dep=1	4373		

Question 4 *(2 points)* : On s'intéresse au sous échantillon des individus appartenant à la CSP "cadres et cadres supérieurs". On suppose que le revenu moyen annuel de ces individus est de 20000 euros, que le prix moyen de leurs achats est de 30 euros. Calculez pour l'individu moyen de cette CSP, la déformation de la probabilité d'achat imputable à l'action marketing.

Partie II : Modèle Tobit (5 points) Les résultats d'estimation par maximum de vraisemblance du modèle tobit simple censuré sont reportés sur la figure (2). On admet que l'estimateur du maximum de vraisemblance est convergent. On note

$$\underset{(1,K)}{x_i} = \left[1\, r_i\, p_i\, m_i\, csp_i^{(1)}\, csp_i^{(2)}\, csp_i^{(3)}\right]$$

$$\underset{(K,1)}{\Upsilon} = (\theta\, \beta_1\, \beta_2\, \beta_3\, \gamma_1\, \gamma_2\, \gamma_3)$$

Question 1 *(1.5 points)* : On considère l'individu moyen de la CSP des "cadres et cadres supérieurs" ($csp_i^{(2)} = 1$) pour lequel on rappelle que $r_i = 20$ et $p = 30$. Calculez pour cet individu une prévision du niveau de consommation potentiel $E\left(c_i^*/x_i\right)$ en l'absence d'action marketing, puis refaites le même calcul en cas d'action marketing. Quelle est alors la variation de consommation potentielle c_i^* imputable, toutes choses égales par ailleurs, à ce type d'action marketing ?

Question 2 *(1.5 points)* : La société X vous demande d'évaluer (toujours pour le même individu moyen) la variation de la consommation effective imputable à l'action marketing. Vous calculerez pour cela la prévision de la variable dépendante $E\left(c_i/x_i\right)$ avec ou sans action marketing, et vous en déduirez la variation consommation effective imputable à cette action marketing.

Figure 2: Estimation du Modèle Tobit

Dependent Variable: CONSO
Method: ML - Censored Normal (TOBIT)
Date: 12/09/03 Time: 10:39
Sample: 1 10000
Included observations: 10000
Left censoring (value) at zero
Convergence achieved after 10 iterations
Covariance matrix computed using second derivatives

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.492957	0.044944	10.96829	0.0000
R	1.600256	0.000488	3281.586	0.0000
P	-1.198181	0.003742	-320.2157	0.0000
M	0.788603	0.016026	49.20675	0.0000
CSP1	-1.017232	0.018660	-54.51495	0.0000
CSP2	2.771211	0.024125	114.8700	0.0000
CSP3	1.500019	0.022928	65.42398	0.0000
Error Distribution				
SCALE:C(8)	0.501245	0.005315	94.30666	0.0000

R-squared	0.999722	Mean dependent var	9.811582
Adjusted R-squared	0.999722	S.D. dependent var	19.72987
S.E. of regression	0.329098	Akaike info criterion	0.659519
Sum squared resid	1082.187	Schwarz criterion	0.665287
Log likelihood	-3289.594	Hannan-Quinn criter.	0.661471
Avg. log likelihood	-0.328959		
Left censored obs	5627	Right censored obs	0
Uncensored obs	4373	Total obs	10000

Question 3 *(2 points)* : Déterminez l'effet marginal associé au revenu r_i et décomposez cet effet marginal en un effet sur la probabilité de consommer et un effet sur le montant de consommation lorsque cette dernière est observable (McDonald[1] et Moffit 1980). Appliquez cette décomposition à l'individu moyen de la CSP "cadres et cadres supérieurs" en l'absence d'action marketing.

[1] McDonald, J. and R. Moffitt (1980) "The Uses of Tobit Analysis", *Review of Economic and Statistics*, 62, pp. 318-321

Exercice 2 (10 points) : Modèle de déséquilibre

On cherche dans le cadre de cet exercice à construire la log-vraisemblance associée à un modèle de déséquilibre suivant la méthodologie proposée par Nelson et Maddala (1974)[2]. Ce type de modèle se classe dans la famille des modèles à régime inobservable et présentent de fortes similarités avec les modèles Tobit censurés.

Les modèles de déséquilibre sont fondés sur l'idée selon laquelle les prix ne s'ajustent qu'imparfaitement et qu'ils ne peuvent à tout moment équilibrer l'offre et la demande sur le marché étudié (Benassy 1976[3]). On note D_t la demande qui dépend d'un ensemble de facteurs $X_{1,t}$ et S_t l'offre supposée dépendre d'un ensemble de facteurs $X_{2,t}$.

$$D_t = X_{1,t}\beta_1 + \varepsilon_{1,t} \tag{3}$$

$$S_t = X_{2,t}\beta_2 + \varepsilon_{2,t} \tag{4}$$

où $\varepsilon_{1,t}$ et $\varepsilon_{2,t}$ désignent les résidus des deux régimes, $\beta_i \in \mathbb{R}^K$ pour $i = 1, 2$ et où les variables explicatives $X_{1,t}$ et $X_{2,t}$ sont continues et observables. En revanche, on suppose que **l'offre et la demande sur le marché ne sont pas directement observables**. En l'absence d'ajustement des prix, la règle nous permet toutefois de postuler que **la quantité effectivement échangée sur le marché notée Q_t, qui est observable, correspond au minimum de l'offre et la demande**.

$$Q_t = \min(D_t, S_t) \tag{5}$$

Ainsi par exemple, on observe la demande ($Q_t = D_t$) dans le cas d'un régime de demande c'est à dire lorsque $D_t < S_t$. On suppose pour simplifier que le vecteur des résidus $\varepsilon_t = (\varepsilon_{1,t}\, \varepsilon_{2,t})'$ est *i.i.d.* et normalement distribué $N(0, \Omega)$, avec :

$$\Omega = E(\varepsilon_t \varepsilon_t') = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \tag{6}$$

On cherche ici à construire la log-vraisemblance associée à un échantillon de T observations $q = (q_1, .., q_T)$ et $x = (x_1', ..., x_T')'$ avec $x = (x_{1,t}\ x_{2,t})$ pour un ensemble de paramètres $\theta = \left(\beta_1'\, \beta_2'\, \sigma_1^2\, \sigma_2^2\right)'$.

$$\widetilde{L}(\theta) = L(\theta, q, x) = \sum_{t=1}^{T} \log\left[f_{Q_t}(q_t, \theta)\right] \tag{7}$$

où $f_{Q_t}(q_t, \theta)$ désigne la densité marginale de Q_t associée à une observation q_t. A partir de cette vraisemblance, Nelson et Maddala (1974) construisent des estimateurs du maximum de vraisemblance des paramètres β.

Les questions peuvent être traitées indépendamment les unes des autres en utilisant les résultats fournis.

On cherche tout d'abord à construire la densité marginale $f_{Q_t}(q_t, \theta)$, notée $f_{Q_t}(q_t)$ pour simplifier les notations, à la base de la construction de la log-vraisemblance. Pour cela, on admet que :

$$f_{Q_t}(q_t) = f_{Q_t|D_t<S_t}(q_t) + f_{Q_t|S_t<Q_t}(q_t) \tag{8}$$

où $f_{Q_t|D_t<S_t}(q_t)$ désigne la densité marginale de Q_t sous l'hypothèse $D_t < S_t$ et $f_{Q_t|S_t<Q_t}(q_t)$ désigne la densité marginale de Q_t sous l'hypothèse $S_t < D_t$. On note $g_{D_t,S_t}(d_t, s_t)$ la densité jointe de D_t et S_t pour des réalisations respectives d_t et s_t de la demande et de l'offre.

Question 1 *(3 points)* : En remarquant que la densité marginale de la demande D_t s'écrit :

$$f_{D_t}(d_t) = \int_{-\infty}^{\infty} g_{D_t,S_t}(d_t, z)\, dz$$

[2] Nelson, F.D,.et Maddala, G.S., (1974), "Maximum Likelihood Methods for Models of Markets in Disequilibrium", *Econometrica,* 42, No. 6, November, pp. 1013-1030.

[3] Benassy, J.P. (1976), "Macroéconomie et théorie du déséquilibre", Dunod

montrez que la densité marginale de la quantité Q_t lorsque $D_t < S_t$ s'écrit sous la forme :

$$f_{Q_t|D_t<S_t}(q_t) = \int_{q_t=d_t}^{\infty} g_{D_t,S_t}(d_t, z)\, dz \tag{9}$$

et que de façon symétrique :

$$f_{Q_t|S_t<D_t}(q_t) = \int_{q_t=s_t}^{\infty} g_{D_t,S_t}(z, s_t)\, dz \tag{10}$$

Question 2 *(2 points)* : En reprenant la définition de la densité marginale d'un vecteur de variables normales, montrez que la densité jointe $g_{D_t,S_t}(d_t, s_t)$ des variables D_t et S_t s'écrit :

$$g_{D_t,S_t}(d_t, s_t) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{1}{2}\left(\frac{q_t - x'_{1,t}\beta_1}{\sigma_1}\right)^2\right] \times \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{1}{2}\left(\frac{s_t - x'_{2,t}\beta_2}{\sigma_2}\right)^2\right] \tag{11}$$

Question 3 *(3 points)* : A partir des résultats des questions 1 et 2, montrez que la densité marginale de la quantité échangée Q_t lorsque l'on est en régime de demande $D_t < S_t$ peut s'écrire sous la forme :

$$f_{Q_t|D_t<S_t}(q_t) = \frac{1}{\sigma_1}\phi\left(\frac{q_t - x'_{1,t}\beta_1}{\sigma_1}\right) \times \frac{1}{\sqrt{2\pi}\sigma_2}\int_{q_t}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{z - x'_{2,t}\beta_2}{\sigma_2}\right)^2\right] dz$$

où $\phi(.)$ désigne la fonction de densité de la loi normale centrée réduite. En posant un changement de variable $\widetilde{z} = \left(z - x'_{2,t}\beta_2\right)/\sigma_2$, avec $dz = d\widetilde{z}\,\sigma_2$, montrez que cette expression se ramène à :

$$f_{Q_t|D_t<S_t}(q_t) = \frac{1}{\sigma_1}\phi\left(\frac{q_t - x'_{1,t}\beta_1}{\sigma_1}\right) \times \Phi\left(\frac{x'_{2,t}\beta_2 - q_t}{\sigma_2}\right) \tag{12}$$

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

Question 4 *(2 points)* : On admet que de façon, symétrique

$$f_{Q_t|S_t<Q_t}(q_t) = \frac{1}{\sigma_2}\phi\left(\frac{q_t - x'_{2,t}\beta_2}{\sigma_2}\right) \times \Phi\left(\frac{x'_{1,t}\beta_1 - q_t}{\sigma_1}\right) \tag{13}$$

Montrez que la log-vraisemblance du modèle de déséquilibre s'écrit alors :

$$\widetilde{L}(\theta) = \sum_{t=1}^{T} \log\left[\frac{1}{\sigma_1}\phi\left(\frac{q_t - x'_{1,t}\beta_1}{\sigma_1}\right) \times \Phi\left(\frac{x'_{2,t}\beta_2 - q_t}{\sigma_2}\right) + \frac{1}{\sigma_2}\phi\left(\frac{q_t - x'_{2,t}\beta_2}{\sigma_2}\right) \times \Phi\left(\frac{x'_{1,t}\beta_1 - q_t}{\sigma_1}\right)\right]$$

Que devient cette expression lorsque σ_1 tend vers 0 ? Quels problèmes cela pose-t-il en termes d'optimisation numérique lorsque l'on cherche à déterminer l'estimateur du maximum de vraisemblance des paramètres ?