

Université d'Orléans - Master Econométrie et Statistique Appliquée

Econométrie des Variables Qualitatives

Examen Mai 2009. C. Hurlin

Exercice 1 (13 points) : Modèle Probit¹

On considère un modèle dichotomique tel que :

$$y_i = \begin{cases} 1 & \text{si } y_i^* \geq 0 \\ 0 & \text{si } y_i^* < 0 \end{cases}, \quad (1)$$

où la variable latente y_i^* vérifie :

$$y_i^* = \alpha + \beta x_i + \varepsilon_i, \quad (2)$$

où $\varepsilon_i \text{ N.i.d. } (0, 1)$ et où la variable x_i est une variable dichotomique prenant deux valeurs $\{0; 1\}$. On dispose de 100 observations réparties de la façon suivante :

Nombre d'Observations	y_i	x_i
16 observations	$y_i = 1$	$x_i = 1$
26 observations	$y_i = 1$	$x_i = 0$
32 observations	$y_i = 0$	$x_i = 1$
26 observations	$y_i = 0$	$x_i = 0$

Question 1 (1 point) : Déterminez en fonction des paramètres α et β , et de la fonction de répartition de la loi normale centrée réduite, noté $\Phi(\cdot)$, la probabilité d'apparition de l'événement $y_i = 1$, notée $\Pr(y_i = 1)$.

Question 2 (2 points) : Déterminez la fonction de log-vraisemblance de ce modèle probit.

Question 3 (2 points) : Montrez que les réalisations des estimateurs du maximum de vraisemblance (MV) des paramètres α et β valent :

$$\hat{\alpha} = \Phi^{-1}(0.5) = 0, \quad (3)$$

$$\hat{\beta} = \Phi^{-1}(1/3) = -0.43,$$

Remarque : on préconise de considérer un changement de variable dans la log-vraisemblance du modèle en posant $z_1 = \Phi(\alpha + \beta)$ et $z_2 = \Phi(\alpha)$, de déterminer les quantités optimales \hat{z}_1 et \hat{z}_2 qui annulent le gradient, et en déduire enfin les estimateurs du MV.

Question 4 (2 points) : Montrez que l'estimation de l'effet marginal associé à la variable x est égal pour tous les individus à $-1/6$. *Remarque : on rappelle que la variable x_i est dichotomique et ne peut prendre que deux valeurs, i.e. 0 et 1.*

Question 5 (2 points) : Testez l'hypothèse nulle $H_0 : \beta = 0$ en utilisant un test de ratio de vraisemblance.

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

Question 6 (2 point) : Donnez la valeur du R^2 de Mac Fadden de ce modèle.

Question 7 (2 points) : Calculez les probabilités estimées de réalisation de l'événement pour les différents types d'individus de l'échantillon et tracez la ROC curve de ce modèle probit.

Exercice 2 (8 points) : Modèle Logit Conditionnel

On considère une enquête sur les modes de transport reliant trois villes qui a été menée en Australie en 1986 auprès de 210 personnes (Bradley, 1993; Hensher et Greene, 2002). Les personnes interrogées pouvaient choisir entre l'avion, le train, le bus ou la voiture. L'enquête leur demandait pour chaque mode de transport d'indiquer le temps de transfert entre leur domicile et le lieu de départ² (variable tps_transf), le temps de trajet (variable tps_trajet) et le coût du voyage (variable $cout$). Le mode de transport vaut 1 pour l'avion, 2 pour le train, 3 pour le bus et 4 pour la voiture.

Question 1 (0.5 point) : Quel modèle proposeriez vous pour modéliser dans ce cas les choix de moyen des transports des individus? Quels sont ses principaux avantages ?

Question 2 (0.5 point) : En supposant qu'il n'existe pas de données manquantes, combien le fichier de données comporte-t-il d'observations au total (nombre de lignes) ? *Vous justifierez précisément votre réponse.*

Question 3 (1 point) On considère à présent le même modèle en introduisant des constantes spécifiques à chaque mode de transport (sauf pour la voiture) sous la forme suivante :

$$\Pr(y_i = j) = \frac{\exp(x_{i,j}\beta)}{\sum_{k=1}^4 \exp(x_{i,k}\beta)}, \quad \forall j \in \{1, 2, 3, 4\}, \quad (4)$$

où l'index vérifie la relation :

$$\begin{aligned} x_{i,j}\beta &= \beta_1 * \mathbb{I}_{(j=1)} + \beta_2 * \mathbb{I}_{(j=2)} + \beta_3 * \mathbb{I}_{(j=3)} + \beta_4 * tps_trans_{i,j} \\ &\quad + \beta_5 * tps_trajet_{i,j} + \beta_6 * cout_{i,j}, \end{aligned} \quad (5)$$

et où $\mathbb{I}_{(j=z)}$ désigne une variable dummy prenant la valeur 1 si $j = z$, et 0 sinon. Commentez cette spécification. Peut-on analyser le signe des coefficients β_1 , β_2 et β_3 associés aux variables dummies ?

Question 4 (2 points) : A partir des résultats d'estimation du modèle de la question 4 (cf. figure 1) et des caractéristiques individuelles (cf. figure 2), (i) calculez la probabilité que l'individu 2 choisisse les différents modes de transport. (ii) Quelle prévision donne alors le modèle?

Question 5 (2 points) : Calculez pour l'individu 2 l'effet marginal d'une hausse de 1 unité du coût de transport associé à la modalité voiture sur la probabilité qu'il prenne la voiture. Commentez.

Question 6 (2 points) : Calculez la log-vraisemblance du modèle de la question 4 associée à l'individu 2, sachant que cet individu a effectivement choisi la voiture comme mode de transport.

²Ce temps est nul dans le cas de la voiture.

Figure 1
Résultats de l'estimation

Parameter Estimates						
Parameter	DDL	Estimation	Erreur standard	t Value	Approx Pr > t	Gradient
avion	1	4.7399	0.8675	5.46	<.0001	3.43E-10
train	1	3.9532	0.4686	8.44	<.0001	2.8E-10
bus	1	3.3062	0.4583	7.21	<.0001	-609E-12
tps_transf	1	-0.0969	0.0103	-9.37	<.0001	-1.88E-8
tps_trajet	1	-0.003995	0.000849	-4.70	<.0001	-2.08E-7
cout	1	-0.0139	0.006651	-2.09	0.0365	1.643E-8

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

Figure 2
Les 8 premières observations de la table en entrée

ident	mode	choix	tps_transf	tps_trajet	cout	revenu	nbre
1	1	0	69	100	59	35	1
1	2	0	34	372	31	35	1
1	3	0	35	417	25	35	1
1	4	1	0	180	10	35	1
2	1	0	64	68	58	30	2
2	2	0	44	354	31	30	2
2	3	0	53	399	25	30	2
2	4	1	0	255	11	30	2

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