

# Université d'Orléans - Master Econométrie et Statistique Appliquée

## Econométrie des Variables Qualitatives

Examen Mai 2010. C. Hurlin

Exercice 1 (13 points) : Modèle Logit<sup>1</sup>

On considère une variable dichotomique prenant deux modalités (0, 1) et telle que :

$$\Pr(y_i = 1) = \Lambda(\alpha + \beta x_i) \quad (1)$$

où  $\Lambda(\cdot)$  désigne la fonction de répartition de la loi logistique et  $x_i$  désigne 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  $\alpha$  et  $\beta$ , et de la variable explicative  $x_i$ , la probabilité d'apparition de l'événement  $y_i = 1$

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

$$\begin{aligned} \ln L(y; \alpha, \beta) = & -48 \times \ln [1 + \exp(-\alpha - \beta)] - 52 \times \ln [1 + \exp(-\alpha)] \\ & -58 \times \alpha - 32 \times \beta, \end{aligned} \quad (2)$$

**Question 3** (2 points) : Montrez que les réalisations des estimateurs du maximum de vraisemblance (MV) des paramètres  $\alpha$  et  $\beta$  valent :

$$\hat{\alpha} = 0, \quad (3)$$

$$\hat{\beta} = -\ln(2), \quad (4)$$

**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  $H_0 : \beta = 0$  au seuil de risque de 5% en utilisant un test de ratio de vraisemblance.

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

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

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

**Exercice 2** (8 points) : Probit polytomique ordonné<sup>2</sup>

On s'intéresse aux déterminants de la fréquentation  $F$  du cinéma. On observe uniquement si l'individu  $i$  ne va jamais au cinéma ( $F_i = 0$ ), va au cinéma au moins une fois tous les deux mois ( $F_i = 1$ ), moins d'une fois par mois ( $F_i = 2$ ) ou au moins une fois par mois ( $F_i = 3$ ). On estime pour cela un modèle probit polytomique à seuils inconnus tel que :

$$F_i = \begin{cases} 0 & \text{si } s_0 < y_i^* \leq s_1 \\ 1 & \text{si } s_1 < y_i^* \leq s_2 \\ 2 & \text{si } s_2 < y_i^* \leq s_3 \\ 3 & \text{si } s_3 < y_i^* < s_4 \end{cases}, \quad \forall i = 1, \dots, N, \quad (5)$$

où  $s_{j+1} \geq s_j$  et où la variable latente  $y_i^*$  est définie par

$$y_i^* = x_i \beta + \varepsilon_i, \quad (6)$$

avec  $x_i = (x_i^1 \dots x_i^K)$ ,  $\beta = (\beta_1 \dots \beta_K)' \in \mathbb{R}^K$ ,  $\varepsilon_i$  *N.i.d.*  $(0, \sigma_\varepsilon^2)$ . Les variables explicatives retenues sont :

- l'âge et le carré de l'âge (Age2, mesurée en années).
- le diplôme avec six modalités qui sont dans l'ordre : >Bac+2, Bac+2, Bac, BEP-CAP, Brevet, et aucun diplôme (modalité de référence).
- le revenu par tranche. Par exemple, rev\_46000=1 si l'individu gagne entre 37000€ et 46000€ par an, la modalité de référence étant un revenu inférieur ou égal à 27000 euros.
- la zone d'habitation. Par exemple, ville\_10000=1 si l'individu habite une ville de moins de 10000 habitants et ville\_plus une ville de plus de 50000 habitants. La modalité de référence correspond aux villes de 10000 à 50000 habitants.

**Question 1** (1 point) : A partir de la sortie de résultats Stata ci-jointe, donnez une estimation des seuils  $s_i$ ,  $i = \{0, \dots, 4\}$ .

**Question 2** (1 point) : Commentez les résultats. En particulier commentez l'effet de l'âge et interprétez l'effet du diplôme.

**Question 3** (1 point) : Exprimez la probabilité estimée d'aller au cinéma au moins une fois tous les deux mois en fonction des paramètres  $\beta$ ,  $\sigma^2$  et  $s_i$ . Le programme n'estime ni la constante de la régression, ni la variance  $\sigma^2$ , pourquoi ?

**Question 4** (1 point) : Quelle est la probabilité estimée d'aller au cinéma au moins une fois tous les deux mois pour un individu ayant le bac, de 35 ans, habitant une ville de 20000 habitants et gagnant 25000 euros par an ?

<sup>2</sup>D'après le contrôle 2009 du cours de Jean Marc Robin "Econométrie des Variables Qualitatives", Université Paris 1, proposé par Jérôme Lè (PSE).

**Question 5** (2 points) : Supposons que l'on ait seulement comme variable explicative l'âge et son carré, et que l'index soit défini par :

$$x_i \widehat{\beta} = \widehat{\beta}_1 \text{age}_i + \widehat{\beta}_2 (\text{age}_i)^2, \quad (7)$$

Calculez *analytiquement* l'effet de l'âge sur la probabilité d'aller au cinéma au moins une fois tous les deux mois. Donnez la valeur de cet effet marginal pour un individu de 35 ans en supposant  $\widehat{\beta}_1 = -0.0505771$  et  $\widehat{\beta}_2 = 0.0002515$ ,  $\widehat{s}_2 = -0.2379$  et  $\widehat{s}_1 = -1.178$ .

**Question 6** (2 points) : On considère à présent le modèle complet. Définissez l'effet marginal du diplôme du baccalauréat sur la probabilité d'aller au cinéma au moins une fois tous les deux mois. *Remarque* : On notera  $Z_j$ ,  $j = \{0, 1, \dots, 5\}$  la variable représentant les différents niveaux de diplôme avec  $Z_0$  la référence et  $Z_3$  la modalité Bac. On notera  $\gamma\omega_i$  la partie de l'index hors diplôme et  $\pi_j$  le coefficient du diplôme, pour les modalités  $j$ .

**Question 7** (1 point) : Ecrivez la contribution d'un individu à la vraisemblance et donnez l'expression de la vraisemblance de l'échantillon.

```
. xi: oprobit cine age age2 i.dipl rev_37000 rev_46000 rev_68000
rev_plus rural
```

```
ville_10000 ville_plus paris
```

```
i.dipl          _Idipl_1-6          (_Idipl_6 for dipl==CEP
omitted)
```

```
Iteration 0:  log likelihood = -55891.898 Iteration 1:  log
likelihood = -46705.809 Iteration 2:  log likelihood = -46554.566
Iteration 3:  log likelihood = -46554.162 Iteration 4:  log
likelihood = -46554.162
```

```
Ordered probit regression
47340
```

```
Number of obs =
LR chi2(15) = 18675.47
Prob > chi2 = 0.0000
Pseudo R2 =
```

```
Log likelihood = -46554.162
0.1671
```

cine	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-.0505771	.0016529	-30.60	0.000	-.0538166	-.0473375
age2	.0002515	.000017	14.82	0.000	.0002182	.0002847
_Idipl_1	1.05227	.0214621	49.03	0.000	1.010205	1.094335
_Idipl_2	.8588332	.0208044	41.28	0.000	.8180574	.899609
_Idipl_3	.755546	.0189757	39.82	0.000	.7183542	.7927377
_Idipl_4	.6097523	.0215757	28.26	0.000	.5674647	.65204
_Idipl_5	.3695112	.0164095	22.52	0.000	.3373491	.4016733
rev_37000	.1475307	.0160734	9.18	0.000	.1160273	.179034
rev_46000	.252039	.0200421	12.58	0.000	.2127572	.2913209
rev_68000	.3613972	.0234549	15.41	0.000	.3154264	.407368
rev_plus	.4322487	.0355369	12.16	0.000	.3625975	.5018998
rural	-.2334771	.0198013	-11.79	0.000	-.2722869	-.1946673
ville_10000	-.1271862	.0233085	-5.46	0.000	-.17287	-.0815025
ville_plus	.1663336	.0179655	9.26	0.000	.1311219	.2015452
paris	.3062037	.0209635	14.61	0.000	.2651159	.3472914
/cut1	-1.178044	.0407032			-1.257821	-1.098267
/cut2	-.2379098	.0403224			-.3169403	-.1588792
/cut3	.0692706	.0402505			-.0096188	.1481601

Extrait de la matrice de variance covariance (obtenu à partir de la commande vce) :

e(V)	cine	age	age2	_Idipl_1	_Idipl_2	_Idipl_3
cine						
age	2.732e-06					
age2	-2.745e-08	2.880e-10				
_Idipl_1	-1.531e-06	3.026e-08	.00046062			
_Idipl_2	-1.144e-06	2.801e-08	.00018294	.00043282		
_Idipl_3	1.127e-06	2.990e-09	.00016868	.00016118	.00036008	
_Idipl_4	4.084e-06	-2.344e-08	.00016795	.00016247	.00015828	
_Idipl_5	-3.057e-06	4.528e-08	.0001531	.00015234	.00014484	