CHAPITRE 4 : LE MODELE LOGIT CONDITIONNEL DE MCFADDEN

SECTION 1 : Exemple : l'achat de boisson

Base de données : BOISSONS.XLS

N=1822 individus avec 3 choix soit 5466 observations

Variable endogène :

$$\mathbf{CHOIX} = \begin{cases} \mathbf{1} & \mathbf{pepsi} \\ \mathbf{2} & \mathbf{seven-up} \\ \mathbf{3} & \mathbf{coca\ cola} \end{cases}$$

Variables exogènes :

PRIX = prix des bouteilles de soda en $

PROMOTION = si une bouteille de soda est en promotion (=1), si non (=0)

AFFICHER = si une bouteille de soda est mis en évidence (=1), si non (=0)

Il y a aussi la variable **ID** pour repérer l'individu. Ce n'est pas une variable exogène.

NB : dans un premier temps on n'utilisera qu'une seule variable exogène : le **PRIX**.

boissons observations

Obs.	ID	CHOIX	PRIX	PROMO	AFFICHER
1	1	0	1.79	0	0
2	1	0	1.79	0	0
3	1	1	1.79	0	0
4	2	0	1.79	0	0
5	2	0	1.79	0	0
6	2	1	0.89	1	1
7	3	0	1.41	0	0
8	3	0	0.84	0	1
9	3	1	0.89	1	0
10	4	0	1.79	0	0
11	4	0	1.79	0	0
12	4	1	1.33	1	0

SECTION 2 : ECRITURE DU MODELE LOGIT CONDITIONNEL

Réécrivons les probabilités du chapitre 3 en incorporant l'indice i sans normaliser le premier vecteur à zéro :

$$\boxed{(4.1)}\quad \boldsymbol{P_{ij} = Proba(y_i = j) = \frac{\exp(cste_{ij} + b_{ij}x_i)}{\sum_{k=1}^{m} \exp(cste_{ik} + b_{ik}x_i)}}$$

On suppose donc dans (4.1) des paramètres différents en fonction des alternatives j i.e. (b_{ij}) et des variables explicatives constantes sur les alternatives j i.e. (x_i).

Dans un <u>modèle conditionnel </u>on considère un vecteur de paramètres constants (b) et on autorise les variables exogènes à dépendre des modalités (x_{ij}) :

(4.2)

$$\boldsymbol{P_{ij} = Proba(y_i = j) = \frac{\exp(cste_{ij} + bx_{ij})}{\sum_{k=1}^{m} \exp(cste_{ik} + bx_{ik})}}$$

On estime toutes ces probabilités par la méthode du maximum de vraisemblance. On en déduit normalement les effets marginaux mais ici on va faire la différence entre les effets marginaux propres (k=j) :

$$\boldsymbol{\frac{dp_{ij}}{dx_{ij}} = \mathrm{p}_{ij}(1 - \mathrm{p}_{ij})b}$$

Et les effets marginaux croisés (k≠j) :

$$\boldsymbol{\frac{dp_{ij}}{dx_{ij}} = -\mathrm{p}_{ij}p_{ik}b}$$

Nous pouvons aussi calculer des ratios de probabilité appelés ratios de risques ou *odds ratios* :

$$\boldsymbol{\frac{p_{ij}}{p_{ik}} = \frac{\exp(cste_{ij} + bx_{ij})}{\exp(cste_{ik} + bx_{ik})} = \exp[(cste_{ij} - cste_{ik}) + b(x_{ij} - x_{ik})]}$$

On remarque que ces *odds ratios* ne dépendent pas du nombre total d'alternatives, ils sont donc indépendants des autres alternatives car ils dépendent des différences de X mais pas de X eux-mêmes.

Cela est dû à l’hypothèse IIA.

SECTION 3 : L’HYPOTHESE IIA

On appelle la propriètè du modèle logit où les *odds ratios* sont indépendants des probabilités restantes l’Indépendance des Alternatives non Pertinentes (Independence from Irrelevant Alternatives (IIA)). Concrétement le rapport des probabilités associées au choix entre deux modalités est indépendant des autres modalités i.e ajouter ou éliminer une troisième modalité ne change pas le rapport entre ces probabilités.

Cette propriété provient de l’hypothèse d’indépendance et d’homoscédasticité des résidus entre les différentes modalités.

NB : l’hypothèse IIA n’est pas pertinente lorsque les modalités sont très similaires (tram, métro), elle est pertinente quand les modalités sont peu comparables (métro, taxi).

Les modèles **probit multinomial**, **nested logit**, **mixed logit** ne nécessitent pas l’hypothèse IIA mais ils sont largement plus difficiles.

Hausman et McFadden (1984) ont construit un test pour vérifier la validité de l'hypothèse IIA. On estime d'abord le modèle complet puis un modèle restreint.

$$\begin{cases} H_0: & IIA\ valide \\ H_1: & IIA\ non\ valide \end{cases}$$

Ce test suit une χ^2 à k degrès de liberté.
Si Pvalue > 5% on accepte H_0.

SECTION 4 : APPLICATION ECONOMETRIQUE

On adapte les expressions (4.2) à la base de données de la section 1. On doit donc estimer :

$$\boldsymbol{P_{i1}} = \frac{\exp(cste_{i1} + b\,prix_{i1})}{\exp(cste_{i1} + b\,prix_{i1}) + \exp(cste_{i2} + b\,prix_{i2}) + exp(b\,prix_{i3})}$$

$$\boldsymbol{P_{i2}} = \frac{\exp(cste_{i2} + b\,prix_{i2})}{\exp(cste_{i1} + b\,prix_{i1}) + \exp(cste_{i2} + b\,prix_{i2}) + exp(b\,prix_{i3})}$$

$$\boldsymbol{P_{i3}} = \frac{\exp(b\,prix_{i3})}{\exp(cste_{i1} + b\,prix_{i1}) + \exp(cste_{i2} + b\,prix_{i2}) + exp(b\,prix_{i3})}$$

Remarque : $cste_{i3}$ = 0 sinon il y aurait multicolinéarité.

Profil réponse discrète			
Index	CHOICE	Fréquence	Pourcentage
0	1	630	34.58
1	2	682	37.43
2	3	510	27.99

Mesures du critère qualificatif de lissage		
Mesure	Valeur	Formule
Likelihood Ratio (R)	354.22	2 * (LogL - LogL0)
Upper Bound of R (U)	4003.3	- 2 * LogL0
Aldrich-Nelson	0.1628	R / (R+N)
Cragg-Uhler 1	0.1767	1 - exp(-R/N)
Cragg-Uhler 2	0.1988	(1-exp(-R/N)) / (1-exp(-U/N))
Estrella	0.1842	1 - (1-R/U)^(U/N)
Adjusted Estrella	0.1812	1 - ((LogL-K)/LogL0)^(-2/N*LogL0)
McFadden's LRI	0.0885	R / U
Veall-Zimmermann	0.2368	(R * (U+N)) / (U * (R+N))
N = # d'observations, K = # de régresseurs		

boissons conditional logit

The MDC Procedure

Conditional Logit Estimates

Résultats estimés des paramètres					
Paramètre	DDL	Valeur estimée	Erreur type	Valeur du test t	Approx. de Pr > |t|
PRIX	1	-2.2964	0.1377	-16.68	<.0001
pepsi	1	0.2832	0.0624	4.54	<.0001
sevenup	1	0.1038	0.0625	1.66	0.0965

b = -2.2964

cste_{i1} = .2832 pour pepsi

cste_{i2} = .1038 pour seven-up

On en déduit les probabilités au point (1 ; 1.25 ; 1.10) respectivement pour Pepsi, 7-up et Coca :

Probabilités estimées

Obs.	ID	phat	PRIX	alt
5467	9998	0.48319	1	1
5468	9998	0.22746	1.25	2
5469	9998	0.28934	1.1	3

Si par exemple on augmente le prix de Pepsi de 1 à 1.10 (les autres restants constants), les probabilités deviennent :

Probabilités estimées

Obs.	ID	phat	PRIX	alt
5467	9998	0.48319	1	1
5468	9998	0.22746	1.25	2
5469	9998	0.28934	1.1	3
5473	9998	0.42632	1.1	
5474	9998	0.25250	1.25	2
5475	9998	0.32119	1.1	3

On peut aussi calculer les effets marginaux propres et croisée (au point initial) :

Effets marginaux propres

Obs.	ID	d11	d22	d33
1	9998	-0.57344	-0.40353	-0.47219

Effets marginaux croisés

Obs.	ID	d12	d13	d23
1	9998	0.25239	0.32105	0.15114

Effets marginaux propres et croisés

	pepsi	7-up	coca
pepsi	-.57344	.25239	.32105
7-up	.25239	-.40353	.15114
coca	.32105	.15114	-.47219

En TD on calculera le test d'Hausman. Donnons ici seulement le résultat et son interprétation.

P-value = .0029742

Puisque P-value < 5% on accepte H1. L'hypothèse IIA est **NON valide**. C'est-à-dire les *odds ratios* ne sont pas indépendants des probabilités restantes i.e éliminer la modalité COCA change le rapport entre ces probabilités. Ce qui est normal ici car les modalités sont trop similaires.