CHAPITRE 7 : LES MODELES A VARIABLES LIMITEES, TRONQUEES-(~~LES MODELES D'AUTOSELECTION~~)

PARTIE I : Les modèles à variables limitées (tronquées)

SECTION 1 : Définitions

⟹ Les données d'un échantillon sont dit **tronquées** lorsque les observations sont faites seulement pour certains individus, constituant un sous-ensemble de la population observée.

Ex : - On cherche à estimer le revenu moyen avec seulement les observations pour la classe des 2% les plus riches. Le degré de troncature est de 98%.

⟹ Les données d'un échantillon sont dit **censurées** lorsqu'une part très importante des observations de la variable endogène prenne une valeur limite.

Ex : - Les achats de biens durables par les ménages [Cf. Tobin (1958)],

- Le nombre d'heures travaillées par les femmes mariées, (la censure est 0)
- Le nombre de billets demandés pour un spectacle (la censure est la capacité de la salle)…

SECTION 2 : Le modèle de régression censurée (le modèle Tobit)

On veut par exemple estimer le modèle suivant appelé modèle à variable latente :

$$\boldsymbol{y^* = cste + bX + \varepsilon}$$

Avec :

ε suivent une loi normale (m,σ)
et

$$y = \begin{cases} 0 & si\ y^* \leq 0 \\ y^* & si\ y^* > 0 \end{cases}$$

On remarque que ce modèle est censuré avec une limite inférieure de zéro (par exemple).

Méthode d'estimation :

Si la variable endogène est censurée (avec une limite inférieure ou supérieure), alors on peut démontrer que les OLS sont biaisés et inconsistants. On va alors utiliser une procédure inventée par TOBIN : la procédure TOBIT qui utilise la méthode du maximum de vraisemblance en utilisant 2 sortes de données : des données non limitées et des données limitées. [Cf. Tobin 1958 Econometrica]

On en déduit les probabilités (si on utilise un PROBIT) :

$$\boldsymbol{P(y=0)=P(y^{*}\leq 0)=\emptyset[(cste+bx)/\sigma]}$$
$$\boldsymbol{P(y>0)=P(y=y^{*})=1-\emptyset[(cste+bx)/\sigma]}$$

$\emptyset$ est la fonction de probabilité de la loi normale.

<u>Remarque</u>

Le résultat précédent fait référence à la densité d'une variable aléatoire <u>tronquée</u>.

Si x a une distribution normale avec une moyenne m et un écart-type σ et une troncature à a alors :

$$\boldsymbol{Prob(x>a)=1-\emptyset\left(\frac{a-m}{\sigma}\right)}$$

On en déduit :

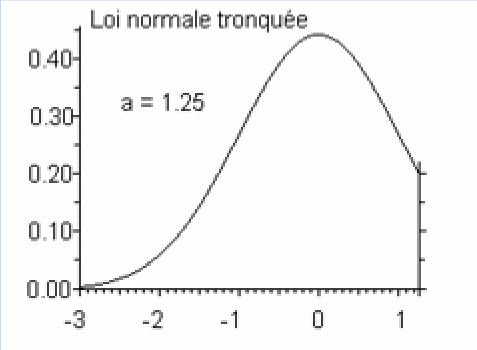
La moyenne tronquée : $\boldsymbol{E(x\,/\,a)=m+\sigma\lambda(\alpha)}$
La variance tronquée : $\boldsymbol{Var(x\,/\,a)=\sigma^{2}[1-\delta(\alpha)]}$

Avec :

$$\alpha=(a-m)/\sigma$$
$$\lambda(\alpha)=\phi(\alpha)/(1-\varphi(\alpha))\quad si\ x>a$$
$$\lambda(\alpha)=-\frac{\phi(\alpha)}{\varphi(\alpha)}\quad si\ x<a$$
$$\delta(\alpha)=\lambda(\alpha)[\lambda(\alpha)-\alpha]$$

φ est la fonction probit.

$\lambda(\alpha)$ est appelé le ratio inverse de Mills (Cf. fonction de hasard)



Mais ce qui nous intéresse réellement ce sont les effets marginaux :

$$\frac{\partial y}{\partial x} = b\varphi\left(\frac{cste + bx}{\sigma}\right)$$

A comparer avec le chapitre 1 :

$$\frac{\partial y}{\partial x} = b\varphi(cste + bx)$$

SECTION 3 : Un exemple : le nombre d'heures travaillées par les femmes mariées

Base de données : TOBIN.XLS

N=753 individus

Variable endogène :

HEURE : Nb d'heures de travail de l'épouse en 2005

Particularité : sur 753 données 325 prennent la valeur 0

Variables exogènes :

REV Revenu imposable du ménage
IFR Impôts fédéraux sur le revenu
MFS Nb de frères et sœurs du mari
FES Nb de frères et sœurs de l'épouse
SPM niveau de scolarité du père du mari
SMM niveau de scolarité de la mère du mari
T2005 1 si l'épouse a travaillé en 2005, sinon 0
ENF6 Nb d'enfants de moins de 6 ans
ENF618 Nb d'enfants de 6 à 18 ans dans le ménage
AGEE âge de l'épouse
SE niveau de scolarité de l'épouse, en années
SALF Salaire horaire moyen de l'épouse de 2005
HEUREM Nb d'heures de travail du mari en 2005
AGEM âge du mari
SM niveau de scolarité de l'époux, en années
SALM salaire du mari en dollars (de 2005)
REVE revenu du ménage en dollars (de 2005)
TAXEE taux d'imposition marginal de l'épouse
SME niveau de scolarité de la mère de l'épouse
SPE niveau de scolarité du père de l'épouse
TCHO Taux de chômage dans la ville de résidence
VILLEGRD 1 si le ménage habite dans une grande ville, 0 sinon
EXPER nb d'années d'expérience sur le marché du travail de l'épouse

On veut estimer :

HEURE = cste +a1 X1 + a2 X2 +……….+ ε

Comment estimer cette équation ?

1 par OLS sur toutes les observations,
2 par OLS sur la partie > 0,
3 par le modèle TOBIT ?

Les résultats globaux se trouvent dans le programme TOBIN.SAS. Pour l'instant donnons les estimations pour un modèle restreint :

$$\widehat{HEURE} = 1349.88 + 73.29\, SE + 80.54\, EXPER - 60.77\, AGEE - 918.92\, ENF6$$

Avec $\hat{\sigma} = 1133.70$

Pour calculer par exemple uniquement l'effet marginal du niveau de scolarité (SE) sur le nombre d'heures travaillées par l'épouse, nous devons choisir un point précis. Nous allons prendre la moyenne des données sauf pour le nombre d'enfants, soit :

12.29	pour SE
10.63	pour EXPER
42.54	pour AGEE
1	pour ENF6

On obtient :

$$\frac{\partial y}{\partial SE} = 73.29\varphi\left(\frac{1349.88 + 73.29 * 12.29 + 80.54 * 10.63 - 60.77 * 42.54 - 918.92 * 1}{1133.70}\right)$$

$$\frac{\partial y}{\partial SE} = 73.29\varphi(-.3504)$$

$$\frac{\partial y}{\partial SE} = 73.29 * .363 = 26.60$$

[NB : pour le calcul de φ voir le site : http://calculis.net/probabilite-loi-normale
~~(NB : pour 'valeur de a' et 'valeur de b', mettre le même nombre i.e. -0.3504)]~~

Interprétation :

Si la scolarité de l'épouse augmente d'une année, cela augmentera le nombre d'heures travaillées, sur une année d'environ 27h, toutes choses égales par ailleurs.

Globalement nous pouvons calculer les probabilités suivantes [Cf. (1.2) du chapitre1] :

$$P(y = 0) = P(y^* \leq 0) = \emptyset[-.3504] = .3751$$

On en déduit :

$$P(y > 0) = P(y = y^*) = .6249$$

Il a 37% de chance pour que l'épouse ne travaille pas et 62% de chance pour qu'elle travaille avec 12.29 pour SE 10.63 pour EXPER, 42.54 pour AGEE, 1 pour ENF6.