MASTER ECONOMETRIE ET STATISTIQUE APPLIQUEE (ESA)

Université d'Orléans

Econométrie des Variables Qualitatives

Chapitre 3

Modèles à Variable Dépendante Limitée

Modèles Tobit Simples et Tobit Généralisés

Christophe Hurlin

Polycopié de Cours

Master Econométrie et Statistique Appliquée (ESA)
Université d'Orléans
Faculté de Droit, d'Economie et de Gestion
Bureau A 224
Rue de Blois – BP 6739
45067 Orléans Cedex 2
www.univ-orleans.fr/deg/masters/ESA/

Contents

Introduction

Nous allons à présent envisager le cas des **modèles à variable dépendante limitée :ce sont des modèles pour lesquels la variable dépendante est continue mais n'est observable que sur un certain intervalle**. Ainsi, ce sont des modèles qui se situent à mi chemin entre les modèles de régression linéaires où la variable endogène est continue et observable et les modèles qualitatifs. En effet, les modèles à variable dépendante limitée dérivent des modèles à variables qualitatives, dans le sens où l'on doit modéliser la probabilité que la variable dépendante appartienne à l'intervalle pour lequel elle est observable. Nous verrons que la structure de base des modèles à variable dépendante limitée est représentée par **le modèle Tobit**. Avant de présenter plus en détail les modèles à variable dépendante limitée, et plus spécifiquement le modèle Tobit, il convient au préalable de préciser les termes que nous allons utilisés par la suite dans le cadre de ce chapitre.

Les modèle Tobit se réfèrent de façon générale à des modèles de régressions dans lesquels le domaine de définition de la variable dépendante est contraint sous une forme ou une autre. En économie, de tels modèles ont été initiés par James Tobin (1958). Son analyse portait sur les dépenses de consommation en biens durables et reposait sur une régression tenant compte spécifiquement du fait que ces dépenses ne peuvent pas être négatives. La variable dépendante était ainsi assujettie à une contrainte de non négativité. Tobin qualifia son modèle **de modèle à variable dépendante limitée**[1] (*limited dependent variables model*) d'où le titre de ce chapitre. Ce modèle et ses généralisations sont plus connus parmi les économistes sous le nom de **modèle Tobit**. Ce terme a été introduit par Goldberger (1964) en raison des similarités avec le modèle probit.

Toutefois, ces modèles sont aussi appelés **modèles de régression censurées** (*censored regression models*) ou **modèle de régression tronquée** (*truncated regression models*). Cette terminologie plus précise permet en effet d'introduire la distinction entre des échantillons tronqués et des échantillons censurés :

1. *Un modèle de régression est dit tronqué lorsque toutes les observations des variables explicatuves et de la variable dépendante figurant en dehors d'un certain intervalle sont totalement perdues.*

2. *Un modèle de régression est dit censuré lorsque l'on dispose au moins des observations des variables explicatives sur l'ensemble de l'échantillon.*

Nous verrons par la suite que **le modèle Tobit est ainsi un modèle de régression censurée**. Les modèles censurés et tronqués ont été utilisés dans d'autres disciplines indépendamment de leur utilisation et développement en économie, et ce notamment en biologie et dans

[1] Tobin J. (1958), "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, 26, 24-36.

les sciences de l'ingénieur. En biologie, de tels modèles furent utilisés pour représenter le temps de survie des patients en fonction de certaines caractéristiques : les échantillons étaient en effet censurés ou tronquées dès lors que le patient reste en vie à la dernière date d'observation de l'échantillon ou si il ne peut pas être ausculté à cette date pour une raison quelconque. De la même façon en ingénierie, les modèles censurés et tronqués sont utilisés pour analyser le temps de survie d'un matériel ou d'un système en fonction de ses caractéristiques. De tels modèles sont alors qualifiés de **modèles de survie** (*survival models*). Les économistes et les sociologues ont aussi utilisés des modèles de survie pour évaluer la durée de phénomènes comme le chômage, le mariage, la durée de résidence dans certains lieux etc... Mathématiquement, les modèles de durée appartiennent à la même classe que les modèles Tobit, mais font souvent l'objet d'un traitement à part.

Entre 1958, date de parution de l'article de Tobin et les années 70, les modèles Tobit ont été utilisés très fréquemment en économie sous l'effet de la conjonction de deux phénomènes : d'une part la plus grande disponibilité de bases micro-économiques et d'autre part le développement des capacités informatiques qui a permis de traiter des modèles Tobit de grande taille. Du fait de ces très nombreuses applications, différentes extensions et généralisations ont été proposées pour le Tobit : **modèle Tobit généralisé**, **modèles à seuils stochastiques**... C'est pourquoi on a introduit la caractérisation de modèle **Tobit simple** pour désigner le modèle développé par Tobin et le distinguer des autres extensions. **Amemiya (1983) identifie ainsi 5 types de modèle de Tobit, le Tobit simple étant qualifié de modèle Tobit Type I**.

Plus formellement, considérons N couples de variables (x_i, y_i^*) où la variable y_i^* est engendrée par un processus aléatoire tel que $E(y_i^*/x_i) = x_i\beta$, où $\beta \in \mathbb{R}^K$ est un vecteur de paramètres. On suppose que la variable y_i^* n'est pas toujours observable : on ne l'observe que si sa valeur est supérieure à un certain seuil c_i. On peut ainsi construire une variable y_i, qui est égale à y_i^* lorsque celle-ci est observable et qui vaut c_i par convention lorsque y_i^* n'est pas observable.

$$y_i = \begin{cases} y_i^* & \text{si } y_i^* > c_i \\ c_i & \text{sinon} \end{cases} \qquad \forall i = 1,..N \tag{0.1}$$

La constante c_i peut être identique pour tous les individus. Deux cas peuvent alors se présenter suivant la nature des observations :

1. Si le vecteur x_i est observable pour tous les individus et cela indépendamment du fait que la variable y_i^* soit observable ou non, on un **échantillon censuré**. Seule la variable y_i^* est observable sur un intervalle $[c_i, +\infty[$

2. Si le vecteur x_i est observable uniquement pour les individus pour lesquels la variable y_i^* est observable, on un **échantillon tronqué**. On ne dispose d'observations (x_i, y_i^*) que pour les individus pour lesquels $y_i^* > c_i$.

On a par exemple un échantillon tronqué dans le cadre d'une enquête où les ménages ne répondent à l'enquête que s'ils répondent à la question permettant de déterminer y_i^*. Ceux pour lesquels $y_i^* \leq c_i$ ne répondent pas à l'enquête ou sont éliminés de l'échantillon par les enquêteurs.

L'utilisation de modèles Tobit suppose que soient particulièrement connus les résultats relatifs aux moments et aux moments conditionnels d'une variable distribuée selon une loi normale tronquée. C'est pourquoi, avant de présenter ces modèles, nous proposons les résultats suivants. Soit $\Phi(.)$ la fonction de répartition de la loi normale centrée réduite $N(0,1)$ et soit $\phi(.)$ la fonction de densité associée.

Proposition 0.1. ***Considérons une variable*** y ***suivant une loi normale tronquée telle que :***

$$y = \begin{cases} y^* & \text{si } y^* > 0 \\ 0 & \text{sinon} \end{cases} \tag{0.2}$$

où y^* ***est distribuée selon une loi normale*** $N(m, \sigma^2)$. ***On admet alors les propriétés suivantes :***

1. ***Espérance de*** y ***:***

$$E(y) = m\,\Phi\left(\frac{m}{\sigma}\right) + \sigma\,\phi\left(\frac{m}{\sigma}\right) \tag{0.3}$$

2. ***Espérance conditionnelle de*** y ***:***

$$E(y/y>0) = m + \sigma\,\frac{\phi\left(\frac{m}{\sigma}\right)}{\Phi\left(\frac{m}{\sigma}\right)} = m + \sigma\,\lambda\left(\frac{m}{\sigma}\right) \tag{0.4}$$

où $\lambda(x) = \phi(x)/\Phi(x)$ ***désigne le ratio de Mill.***

3. ***Variance de*** y ***:***

$$V(y) = \sigma^2\left[\Phi\left(\frac{m}{\sigma}\right) + \frac{m}{\sigma}W\left(\frac{m}{\sigma}\right) - W^2\left(\frac{m}{\sigma}\right)\right] \tag{0.5}$$

avec $W(x) = \int_{-\infty}^{x}\Phi(t)\,dt = x\Phi(x) + \phi(x)$. ***Par conséquent :***

$$\begin{aligned} V(y) \quad = \quad & m^2\Phi\left(\frac{m}{\sigma}\right) + m\,\sigma\,\phi\left(\frac{m}{\sigma}\right) + \sigma^2\,\Phi\left(\frac{m}{\sigma}\right) - m^2\,\Phi^2\left(\frac{m}{\sigma}\right) \\ & -2\,m\,\sigma\,\phi\left(\frac{m}{\sigma}\right)\Phi\left(\frac{m}{\sigma}\right) - \sigma^2\,\phi^2\left(\frac{m}{\sigma}\right) \end{aligned}$$

4. ***Variance conditionnelle de*** y ***:***

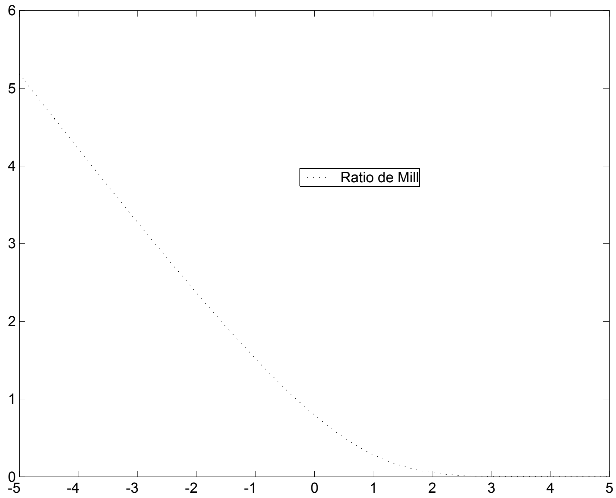
$$V(y/y>0) = \sigma^2\left[1 - \frac{m}{\sigma}\lambda\left(\frac{m}{\sigma}\right) - \lambda^2\left(\frac{m}{\sigma}\right)\right] \tag{0.6}$$

Notons simplement que puisque le ratio de Mill $\lambda(x)$ joue un grand rôle dans l'analyse des moments d'une loi normale tronquée, il est intéressant de vérifier qu'il s'agit d'une fonction décroissante de x :

$$\frac{\partial\lambda(x)}{\partial x} = \frac{-\lambda(x)\,W(x)}{\Phi(x)} \tag{0.7}$$

La forme générale du ratio de Mill est reproduite sur la figure (0.1).

Figure 0.1: Ratio de Mill : $\lambda(x) = \phi(x) / \Phi(x)$

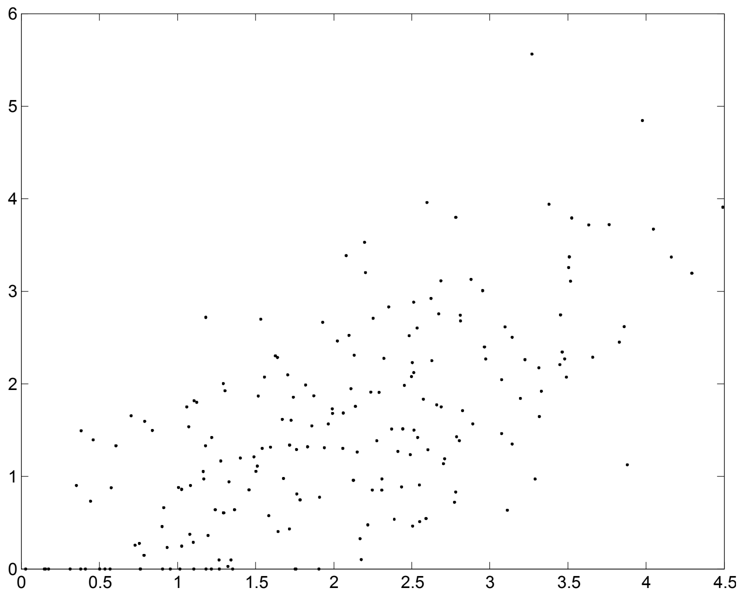


Etudions à présent le modèle Tobit Simple ou modèle Tobit de type 1 suivant la terminologie d'Amemiya (1983).

1. Le Modèle Tobit Simple

Comme nous l'avons dit en introduction le modèle Tobit[2] a été développé par Tobin (1958), même si le terme de modèle Tobit n'est apparu qu'en 1964 dans un article de Goldberger. Dans son étude, Tobin cherche à modéliser la relation entre le revenu d'un ménage et les dépenses en biens durables. Il dispose pour cela d'un échantillon de $N = 735$ consommateurs tiré du *Survey of Consumer Finances*. Tobin observe que lorsque l'on représente les couples revenus - dépenses des N consommateurs, la relation obtenue ressemble au graphique (1.1) ci-dessous. Une des caractéristiques essentielles des données étant que **plusieurs observations pour le montant des dépenses de consommation sont nulles**. En effet, ces observations sont nulles pour tous les ménages n'ayant pas acheté de biens durables sur la période. Pour ces individus, on dispose ainsi d'observations sur le revenu mais pas d'observations sur les dépenses de consommation : on un échantillon censuré.

Figure 1.1: Nuage de Points : Modèle Tobit Simple



Cette propriété remet en cause l'hypothèse de linéarité et montre que les moindres carrés ordinaires ne sont pas la méthode pertinente pour estimer une telle relation. De façon plus générale, **on peut pas ici utiliser une densité continue pour expliquer la distribution conditionnelle des dépenses par rapport au revenu** : en effet, une distribution continue est incompatible avec le fait que plusieurs observations des dépenses soient nulles. C'est donc dans ce contexte que Tobin propose son modèle à variable dépendante limitée (*limited dependent variable model*).

[2] Dans cette section nous parlerons de modèle Tobit en réference au Tobit simple pour alléger les notations.

L'histoire que propose Tobin est alors la suivante. Considérons un agent qui a le choix entre deux biens x et y, qui cherche à maximiser son utilité $U(x,y)$ sous sa contrainte de budget de la forme $x + py \leq R$, où p est le prix relatif et R le revenu. On suppose que le prix du bien x sert de numéraire. On admet parallèlement que la consommation de bien x satisfait une contrainte de non négativité $x \geq 0$, mais que la consommation de bien y vérifie une contrainte du type $y \geq y_0$ ou $y = 0$. Cette contrainte traduit simplement une indivisibilité des premières unités de biens y. Supposons que y^* soit la solution du programme de maximisation de l'utilité sous la contrainte de budget et la contrainte $x \geq 0$.

$$\begin{array}{rcl} (x^*, y^*) & = & \arg\max\limits_{\{x,y\}} U(x,y) \\ sc & : & x + py \leq R. \\ sc & : & x \geq 0 \end{array}$$

Dès lors, deux cas sont à considérer : soit le niveau de consommation potentielle du bien y^* est suffisamment élevé par rapport au seuil y_0 et l'agent consomme effectivement du bien y en quantité y^*, soit il n'est pas suffisamment élevé et l'agent ne consomme pas de bien y. Formellement on a :

$$y = \left\{ \begin{array}{ll} y^* & \text{si } y^* > y_0 \\ 0 & \end{array} \right.$$

Si l'on suppose que la solution non contrainte y^* est fonction d'un certains nombres de caractéristiques x et d'une perturbation ε sous la forme $y^* = \beta_0 + \beta_1 x + \varepsilon$ et si l'on suppose la normalité des perturbations ε , alors on peut reproduire des valeurs de la consommation y semblables à celles du graphiques (1.1). Il suffit pour cela de supposer que les seuils y_0 sont les mêmes pour tous les individus et que $y_0 = 0$.

Ainsi, le modèle originellement proposé par Tobin (1958) est le suivant :

Definition 1.1. ***Un modèle Tobit Simple ou modèle Tobit de type I est défini par :***

$$y_i^* = x_i\beta + \varepsilon_i \qquad \forall i = 1, ..N \tag{1.1}$$

$$y_i = \left\{ \begin{array}{ll} y_i^* & si\ y_i^* > 0 \\ 0 & si\ y_i^* \leq 0 \end{array} \right. \tag{1.2}$$

où $x_i = \left(x_i^1..x_i^K\right), \forall i = 1,..,N$ ***désigne un vecteur de caractéristiques observables et où*** $\beta = (\beta_1...\beta_K)' \in \mathbb{R}^K$ ***est un vecteur de paramètres inconnus et où les perturbations*** ε_i ***sont distribués selon une loi*** $N\left(0, \sigma_\varepsilon^2\right)$**.**

On suppose ainsi que les variables y_i et x_i sont observées pour tous les individus, mais que les variables y_i^* sont observables uniquement si elles sont positives. On note X la matrice de dimension (N, K) telles que les lignes de cette matrice correspondent aux vecteurs x_i. On suppose en outre que

$$\lim_{N\to\infty} X'X = Q_X$$

où Q_X est une matrice définie positive.

Remarquons que l'écriture d'un seuil nul $y_i^* > 0$ peut parfaitement être changé en un seuil $y_i^* > y_0$ sans que le modèle soit changé. Il suffit pour cela d'absorber dans le vecteur des caractéristiques x_i une constante et de lui associer un coefficient égal à y_0. Le cas où les seuils $y_{i,0}$ diffèrent selon les individus nécessite toutefois de modifier le modèle.

Essayons à présent de comprendre pourquoi l'application d'une méthode de moindres carrés ordinaires ne permet pas d'estimer de façon convergente le vecteur des paramètres β associés aux variables explicatives.

1.1. Estimation par les Moindres Carrés Ordinaires

Au delà des calculs, on observe immédiatement sur l'exemple de la figure (1.1) que l'application des Moindres Carrés Ordinaires n'est pas la méthode adéquate pour révéler la relation entre consommation et revenu pour au moins eux raisons :

1. Le nuage de point sera alors mal décrit par une relation du type $consommation = a + b * revenu$ puisque le nuage de points comporte deux parties distinctes.
2. L'hypothèse de loi continue généralement faite sur les perturbations n'est pas adaptée dans ce cas puisque la valeur nulle de la consommation est observée de nombreuses fois dans l'échantillon et a donc sans doute une probabilité d'apparition nettement différente de zéro.

Nous allons toutefois montrer l'application des MCO à l'ensemble des observations ou l'application des MCO aux seules observations pour lesquelles on observe la variable y^* conduit à une estimation biaisée des paramètres du vecteur β.

On suppose que les N observations de l'échantillon sont générées à partir du processus générateur de données suivant :

$$y_i = \begin{cases} y_i^* & \text{si } y_i^* = x_i\beta + \varepsilon_i > 0 \\ 0 & \text{sinon} \end{cases} \qquad \forall i = 1, ..N$$

avec $x_i = \left(x_i^1..x_i^K\right), \forall i = 1, .., N$, $\beta = (\beta_1...\beta_K)' \in \mathbb{R}^K$ et où les perturbations ε_i sont distribués selon une loi $N\left(0, \sigma_\varepsilon^2\right)$. On cherche ici à estimer le vecteur de paramètre β par le méthode des Moindres Carrés Ordinaires. Deux solutions sont alors envisageables :

1. Soit on applique les MCO à l'ensemble des observations (x_i, y_i) de l'échantillon
2. Soit on applique les MCO aux seules observations (x_i, y_i) pour lesquels $y_i^* > 0$.

Commençons tout d'abord par appliquer les MCO à l'ensemble des observations de l'échantillon.

1.1.1. Application des MCO à l'ensemble des observations

L'estimateur des MCO appliqué à l'ensemble des N couples d'observations (y_i, x_i) est défini par la relation suivante :

$$\widehat{\beta}_{LS} = \left(\sum_{i=1}^{N} x_i' x_i\right)^{-1} \sum_{i=1}^{N} x_i' y_i \tag{1.3}$$

Supposons pour commencer que les variables exogènes x_i sont déterministes et déterminons alors l'expression de $E\left(\widehat{\beta}_{LS}\right)$, comme suit :

$$E\left(\widehat{\beta}_{LS}\right)=\left(\sum_{i=1}^{N}x_i'x_i\right)^{-1}\sum_{i=1}^{N}x_i'E\left(y_i\right) \tag{1.4}$$

Cette expression dépend de la quantité $E\left(y_i\right)$ qui correspond à l'espérance d'une variable normale tronquée. En appliquant la formule de l'espérance d'une loi normale tronquée, on montre que :

$$E\left(y_i\right)=x_i\beta\,\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)+\sigma_\varepsilon\,\phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)$$

On en déduit immédiatement que l'estimateur des moindres carrés $\widehat{\beta}$ est biaisé : en effet la quantité $E\left(\widehat{\beta}_{LS}\right)$ *est une fonction non linéaire de* β *et ne peut donc pas être égale à* β. Le biais peut être positif ou négatif et pour le caractériser, considérons le cas où $K=1$, on a alors $\widehat{\beta}_{LS}=\left(\sum_{i=1}^{N}x_i^2\right)^{-1}\sum_{i=1}^{N}x_iy_i$ et l'on en déduit que :

$$E\left(\widehat{\beta}_{LS}\right)=\frac{\sum_{i=1}^{N}x_iE\left(y_i\right)}{\sum_{i=1}^{N}x_i^2}=\left(\frac{1}{\sum_{i=1}^{N}x_i^2}\right)\sum_{i=1}^{N}\left[x_i^2\,\beta\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)+x_i\sigma_\varepsilon\,\phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\right]$$

Si l'on admet que

$$\lim_{n\rightarrow\infty}V\left(\widehat{\beta}_{LS}\right)=0$$

alors, on en déduit que l'estimateur $\widehat{\beta}_{LS}$ converge vers $E\left(\widehat{\beta}_{LS}\right)$ qui dans le cas général diffère de la vraie valeur β des paramètres :

$$\widehat{\beta}_{LS}\underset{N\rightarrow\infty}{\overset{p}{\longrightarrow}}\widetilde{\beta}\neq\beta \tag{1.5}$$

L'estimateur des MCO **de** β **appliqué sur l'ensemble des observations est non convergent**. Il est alors relativement difficile dans le cas de variables déterministes de donner un résultat général sur la forme du biais, c'est à dire sur le fait que l'estimateur $\widehat{\beta}_{LS}$ sur-estime ou sous estime la vraie valeur β des paramètres. C'est pourquoi, nous allons à présent envisager le cas de variables explicatives stochastiques.

Envisageons à présent le cas où les variables x_i sont des variables aléatoires. Goldberger (1981) a étudié les biais asymptotiques de l'estimateur des MCO dans ce cas en supposant *que les toutes les variables explicatives* x_i, *à l'exception du terme constant, étaient distribuées selon une loi normale.* Goldberger réécrit ainsi le modèle sous la forme :

$$y_i=\begin{cases}y_i^* & \text{si } y_i^*=\alpha+x_i\beta+\varepsilon_i>0\\ 0 & \text{sinon}\end{cases}\qquad\forall i=1,..N$$

avec $x_i=\left(x_i^1..x_i^K\right)$, $\forall i=1,..,N$, $\beta=\left(\beta_1...\beta_K\right)'\in\mathbb{R}^K$ et $\gamma\in\mathbb{R}$. Les résidus ε_i sont distribués selon une loi normale $N\left(0,\sigma_\varepsilon^2\right)$.

Hypothèse *On suppose que les variables explicatives* x_i *sont distribuées selon une loi normale* $N\left(0,\Omega\right)$ *avec cov* $\left(\varepsilon_ix_i^{(k)}\right)=0$, $\forall k=1,..,K$.

L'hypothèse de nullité de l'espérance des variables explicatives x_i n'est pas gênante ici puisque si l'on considère des variables non centrées, on peut sans problème intégrer cette quantité dans le terme constant α.

Proposition 1.2. ***Sous les hypothèses de Goldberger (1981), l'estimateur*** $\widehat{\beta}_{LS}$ ***des Moindres Carrés Ordinaires obtenu sur l'ensemble des observations*** (x_i, y_i) ***vérifie :***

$$\widehat{\beta}_{LS} \underset{N\rightarrow\infty}{\overset{p}{\longrightarrow}} \beta \times \Phi\left(\frac{\alpha}{\sigma_y}\right) \tag{1.6}$$

où α ***correspond à la constante de l'équation*** $y_i^* = \alpha + x_i\beta + \varepsilon_i$ ***et*** $\sigma_y^2 = \sigma_\varepsilon^2 + \beta'\Omega\beta$, ***où*** Ω ***désigne la matrice de variance covariance des variables explicatives*** x_i.

La démonstration de cette proposition figure dans Greene (1981)

Exemple : Considérons le cas où $\alpha = 0$. Etant donné que $\Phi(0) = 0.5$, lorsque il n'y pas de constante dans la définition de la variable latente y_i^* $(\alpha = 0)$, alors on obtient une relation du type $plim\left(\widehat{\beta}_{LS}\right) = 0.5 \times \beta$. Dans ce cas, l'estimateur obtenu sur la totalité de l'échantillon converge asymptotiquement vers la moitié de la vraie valeur β des paramètres. En effet, sous l'hypothèse de normalité avec $E(x_i) = 0$, si la constante α est nulle, on a alors $E(y_i^*) = 0$. La variable y_i^* est centrée et distribuée selon une loi symétrique, la loi normale $N\left(x_i\beta, \sigma_\varepsilon^2\right)$. Dès lors sous l'hypothèse de Goldberger lorsque $\alpha = 0$, on a $Prob(y_i^* > 0) = Prob(y_i^* \leq 0) = 0.5$. Pour un échantillon de taille N suffisante, on a donc approximativement autant d'observations nulles de y_i que d'observations strictement positives : $N_1 \simeq N/2$. Dès lors, la prise en compte de l'ensemble des observations dans l'estimation des MCO va conduire à un estimateur de β convergeant vers la moitié de la vraie valeur du vecteur β. Dans le cas $K = 1$, la pente de la droite d'ajustement linéaire associée à la régression sur l'ensemble des observations (c'est à dire $\widehat{\beta}_{LS}$) correspond dans ce cas à la moitié de la pente β associée à la vraie relation linéaire entre y_i^* et x_i.

****

**** Insérer Graphique avec $\alpha = 0$ ****

****

Une des conséquences remarquables de cette proposition est la suivante :

Remark 1. ***Sous les hypothèses de Goldberger (1981), l'estimateur défini par la quantité*** $\widehat{\beta}_{LS}^c = (N/N_1) \times \widehat{\beta}_{LS}$, ***où*** N_1 ***est le nombre d'observations pour lesquelles*** $y_i^* > 0$, ***est un estimateur convergent de*** β. ***Un estimateur convergent de*** α ***peut être obtenu de façon similaire.***

$$\widehat{\beta}_{LS}^c = \frac{N}{N_1}\widehat{\beta}_{LS} \underset{N\rightarrow\infty}{\overset{p}{\longrightarrow}} \beta \tag{1.7}$$

De la même façon pour l'estimateur corrigé $\widehat{\alpha}_{LS}^c$ de la constante α, on a :

$$\widehat{\alpha}_{LS}^c = \frac{N}{N_1}\widehat{\alpha}_{LS} \underset{N\rightarrow\infty}{\overset{p}{\longrightarrow}} \alpha \tag{1.8}$$

Reprenons l'exemple du cas où le terme constant est nul : $\alpha = 0$. On a vu alors que l'estimateur des MCO était biaisé et convergeait vers la moitié de la vraie valeur des paramètres :

$$\widehat{\beta}_{LS} \underset{N\rightarrow\infty}{\overset{p}{\longrightarrow}} \beta \times \Phi\left(\frac{0}{\sigma_y}\right) = \frac{\beta}{2}$$

Or sous les hypothèses de Goldberger et en particulier sous l'hypothèse $E(x_i) = 0$, imposer la nullité de α revient à imposer la nullité de $E(y_i^*)$. Comme nous l'avons dit la variable y_i^* est alors centrée et distribuée selon une loi symétrique, la loi normale $N\left(x_i\beta, \sigma_\varepsilon^2\right)$. Dès lors, lorsque $\alpha = 0$, on a $Prob(y_i^* > 0) = Prob(y_i^* \leq 0) = 0.5$. Pour un échantillon de taille N suffisante, on a donc approximativement autant d'observations nulles de y_i que d'observations strictement positives :

$$N_1 \underset{N\rightarrow\infty}{\overset{p}{\longrightarrow}} \frac{N}{2}$$

Ainsi l'estimateur corrigé $\widehat{\beta}_{LS}^c$ est convergent puisque :

$$\widehat{\beta}_{LS}^c = \frac{N}{N_1} \times \widehat{\beta}_{LS} \underset{N\rightarrow\infty}{\overset{p}{\longrightarrow}} 2\,plim\left(\widehat{\beta}_{LS}\right) = \beta$$

Greene (1983) dérive les matrices de variance covariance asymptotiques de cet estimateur. Malheureusement, on peut utiliser cet estimateur que dans la mesure où l'on est sur que les hypothèses de Goldberger sont satisfaites et en particulier l'hypothèses selon laquelle les variables explicatives sont distribuées selon des lois normales. Rien n'est spécifié sur les propriétés de cet estimateur lorsque les variables explicatives ne sont pas distribuées selon une loi normale. Il faudra donc utiliser une autre méthode pour estimer β dans le cas général.

****

*** Illustrer par simulation biais sur $\widehat{\beta}_{LS}$ et convergence de $\widehat{\beta}_{LS}^c$ ***

****

Appliquons à présent la méthode des MCO aux seules observations pour lesquelles $y_i^* > 0$ afin d'estimer le vecteur de paramètres β.

1.1.2. Application des MCO aux observations pour lesquelles $y_i^* > 0$

Compte tenu du graphique (1.1), il était clair que l'application des MCO à l'ensemble des observations (x_i, y_i) de l'échantillon devait conduire à une estimation biaisée du coefficient qui lie le revenu à la consommation. C'est ce que nous avons démontré dans la section précédente. Mais lorsque l'on restreint l'échantillon aux seules observations pour lesquelles la variable latente y_i^* est positive, ce résultat est beaucoup moins évident à illustrer graphiquement.

Appliquons ainsi les MCO aux seules observations pour lesquelles $y_i^* > 0$. L'estimateur des MCO, noté $\widehat{\beta}_{y>0}$, est défini par la relation :

$$\widehat{\beta}_{y>0} = \left(\sum_{y_i>0} x_i'x_i\right)^{-1} \sum_{y_i>0} x_i'y_i \tag{1.9}$$

où $\sum_{y_i>0}$ désigne la sommation sur les indices $i = 1,..,N$ pour lesquels on a $y_i > 0$. Supposons que les variables exogènes x_i sont déterministes et déterminons alors l'expression de $E\left(\widehat{\beta}_{y>0}\right)$, comme suit :

$$E\left(\widehat{\beta}_{y>0}\right) = \left(\sum_{y_i>0} x_i'x_i\right)^{-1} \sum_{y_i>0} x_i'E\left(y_i/y_{i>0}\right) \tag{1.10}$$

Cette expression dépend de la quantité $E\left(y_i/y_{i>0}\right)$ qui correspond à l'espérance conditionnelle d'une variable normale tronquée. En appliquant la formule correspondante, on montre que :

$$E\left(y_i/y_{i>0}\right) = x_i\beta + \sigma_\varepsilon \frac{\phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)}{\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)} = x_i\beta + \sigma_\varepsilon\,\lambda\left(\frac{x_i\beta}{\sigma_\varepsilon}\right) \tag{1.11}$$

où $\lambda(x) = \phi(x)/\Phi(x)$ désigne le ratio de Mill. Dans ce cas encore, l'estimateur des MCO est donc biaisé et le biais peut être positif et négatif :

$$E\left(\widehat{\beta}_{y>0}\right) = \frac{\sum_{y_i>0} x_i E\left(y_i/y_{i>0}\right)}{\sum_{y_i>0} x_i^2} = \beta + \sigma_\varepsilon \left[\frac{\sum_{y_i>0} x_i\lambda\left(x_i\beta/\sigma_\varepsilon\right)}{\sum_{y_i>0} x_i^2}\right] \tag{1.12}$$

Envisageons à présent le cas où les variables x_i sont distribuées selon une loi normale. On se place alors dans le cadre des hypothèses de Goldberger (1981) décrites à la section précédente. *On suppose que les variables explicatives x_i sont distribuées selon une loi normale $N(0,\Omega)$ avec cov* $\left(\varepsilon_i x_i^{(k)}\right) = 0$, $\forall k = 1,..,K$. Sous ces hypothèses, Goldberger obtient le résultat suivant :

Proposition 1.3. ***Sous les hypothèses de Goldberger (1981), l'estimateur $\widehat{\beta}_{y>0}$ des Moindres Carrés Ordinaires obtenu sur les seules observations (x_i, y_i) pour lesquelles $y_i^* > 0$ vérifie :***

$$\widehat{\beta}_{y>0} \underset{N\rightarrow\infty}{\overset{p}{\longrightarrow}} \left(\frac{1-\gamma}{1-\rho^2\gamma}\right)\beta \tag{1.13}$$

les paramètres γ et ρ étant respectivement définis par :

$$\gamma = \frac{1}{\sigma_y}\lambda\left(\frac{\alpha}{\sigma_y}\right)\left[\alpha + \sigma_y\lambda\left(\frac{\alpha}{\sigma_y}\right)\right] \tag{1.14}$$

$$\rho^2 = \frac{1}{\sigma_y^2}\beta'\Omega\beta \tag{1.15}$$

où α correspond à la constante de l'équation $y_i^* = \alpha + x_i\beta + \varepsilon_i$ et où $\sigma_y^2 = \sigma_\varepsilon^2 + \beta'\Omega\beta$, avec Ω matrice de variance covariance des variables explicatives x_i.

La démonstration de cette proposition figure dans Goldberger (1981). On peut montrer que les paramètres γ et ρ vérifient :

$$0 \leq \gamma \leq 1 \quad\quad 0 \leq \rho^2 \leq 1$$

Dés lors, de façon générale *on montre que l'estimateur des MCO appliqué aux seules observations $y_i > 0$ sous estime l'ensemble des composantes du vecteur β.*

$$\underset{N\rightarrow\infty}{plim}\ \widehat{\beta}_{y>0} \leq \beta \tag{1.16}$$

Une des conséquences remarquable de cette proposition est la suivante :

Remark 2. ***Sous l'hypothèse de normalité des variables x_i, le degré de sous estimation est totalement uniforme pour tous les éléments de β.***

$$\frac{1}{\beta^{(k)}}\underset{N\rightarrow\infty}{plim}\ \widehat{\beta}_{y>0}^{(k)} = \xi \quad\quad \forall k = 1,..,K \tag{1.17}$$

Ainsi le biais affecte de façon symétrique l'ensemble des paramètres estimés. Ce résultat n'est plus valable dès lors que l'on lève l'hypothèse de normalité.

*** Déterminer le cas particulier où l'estimateur $\widehat{\beta}_{y>0}$ n'est pas biaisé c'est à dire lorsque

$$\gamma = \sigma_y^{-1}\lambda\left(\frac{\alpha}{\sigma_y}\right)\left[\alpha + \sigma_y\lambda\left(\frac{\alpha}{\sigma_y}\right)\right] = \frac{\alpha}{\sigma_y}\lambda\left(\frac{\alpha}{\sigma_y}\right) + \lambda^2\left(\frac{\alpha}{\sigma_y}\right) = 1$$

****

*** Insérer Simulations Biais + Graphique pour le cas particulier **
****

1.2. Estimation par la méthode en deux étapes : Heckman (1976)

Puisque la méthode des Moindres Carrés Ordinaires ne peut conduire qu'à des estimations biaisées des paramètres dans le cas d'un modèle Tobit simple, sauf dans des cas très particuliers, différentes méthodes d'estimation alternatives ont été proposées. La méthode d'estimation qui est la plus utilisée aujourd'hui est celle du maximum de vraisemblance (Goldberger 1981, Olsen 1978). Toutefois cette méthode est relativement "gourmande" en termes de capacités de calcul, notamment dans la phase d'optimisation. C'est pourquoi, dans les années 70, du fait des contraintes informatiques, d'autres méthodes d'estimation ont souvent été privilégiées parce qu'elles nécessitaient moins de capacités de calcul : tel est le cas de la méthode d'estimation en deux étapes d'Heckman (1976).

Heckman (1976), suivant une suggestion de Gronau (1974), propose un estimateur en deux étapes dans un modèle Tobit généralisé à deux équations (modèle que nous aborderons dans les sections suivantes). Cet estimateur peut aussi être utilisé pour estimer les paramètres d'un modèle Tobit simple ou modèle Tobit de type I. Pour comprendre cette méthode, considérons la formule de l'espérance conditionnelle de y_i sachant que $y_i > 0$:

$$E\left(y_i/y_{i>0}\right) = x_i\beta + \sigma_\varepsilon\,\lambda\left(\frac{x_i\beta}{\sigma_\varepsilon}\right) \tag{1.18}$$

où $\lambda(x) = \phi(x)/\Phi(x)$ désigne le ratio de Mill. Ainsi, l'espérance conditionnelle de y_i sachant $y_i > 0$ peut être décomposée en une composante linéaire en β et une composante non linéaire en β. Considérons à présent la partie quantitative du modèle Tobit, c'est à dire celle qui correspond à l'observation de $y_i > 0$. Pour ces observations on a une relation du type :

$$y_i = E\left(y_i/y_{i>0}\right) + v_i$$

où v_i est de moyenne nulle. On remplace alors l'espérance conditionnelle par son expression, et l'on obtient la relation suivante.

Proposition 1.4. ***Le modèle Tobit simple, pour*** $y_i > 0$, ***peut être représenté par la régression non linéaire hétéroscédastique suivante :***

$$y_i = x_i\beta + \sigma_\varepsilon\,\lambda\left(x_i\delta\right) + v_i \tag{1.19}$$

avec $\delta = \beta/\sigma_\varepsilon$ ***et*** $v_i = y_i - E\left(y_i/y_{i>0}\right)$ ***et*** $E\left(v_i\right) = 0$ ***et***

$$Var\left(v_i\right) = \sigma_\varepsilon^2 - \sigma_\varepsilon^2 x_i\delta\lambda\left(x_i\delta\right) - \sigma_\varepsilon^2\lambda\left(x_i\delta\right)^2 \tag{1.20}$$

Ainsi, pour estimer les paramètres β, il suffit de considérer la régression non linéaire (1.19) et d'en déduire un estimateur $\widehat{\beta}_H$ à partir des N_1 observations pour lesquelles $y_i > 0$. La seule difficulté provenant de l'hétéroscédasticité des perturbations v_i, puisque $Var(v_i)$ dépend des caractéristiques x_i via le ratio de Mill $\lambda(x_i\delta)$ et directement dans l'expression $x_i\delta\lambda(x_i\delta)$. Telle est l'idée de la procédure d'Heckman.

Proposition 1.5. ***La procédure d'estimation d'Heckman (1976) comporte deux étapes :***

1. ***Etape 1 : Estimer le ratio $\delta = \beta/\sigma_\varepsilon$ à partir du modèle probit dichotomique suivant par une méthode de maximum de vraisemblance***

$$z_i = \begin{cases} 1 & si\ y_i > 0 \\ 0 & sinon \end{cases} \qquad \forall i = 1, ..N \tag{1.21}$$

avec $Prob(z_i = 1) = \Phi(x_i\beta/\sigma_\varepsilon) = \Phi(x_i\delta)$. Soit $\widehat{\delta}$ l'estimateur du MV de δ.

2. ***Etape 2 : Régresser y_i sur x_i et $\lambda\left(x_i\widehat{\delta}\right)$ par une méthode de Moindres Carrés en ne considérant uniquement les N_1 valeurs positives de y_i***

$$y_i = x_i\widehat{\beta}_H + \widehat{\sigma}_\varepsilon\,\lambda\left(x_i\widehat{\delta}\right) + v_i \tag{1.22}$$

On note alors $\widehat{\gamma}_H = \left(\widehat{\beta}_H'\ \widehat{\sigma}_\varepsilon\right)'$ l'estimateur des paramètres du modèle Tobit ainsi obtenu.

En effet sous la forme (1.19), le modèle quantitatif apparaît comme un modèle linéaire en β et σ_ε.

Réécrivons le modèle sous forme vectorielle pour dériver les lois asymptotiques. On pose $\widehat{Z} = \left(X\ \widehat{\lambda}\right)$ où X désigne la matrice de dimension (N_1, K) dont les lignes correspondent aux vecteur de variables explicatives x_i pour lesquelles $y_i > 0$ et où $\widehat{\lambda}$ désigne un vecteur de dimension $(N_1, 1)$ dont le $j^{ème}$ élément est donné par l'estimateur du ratio de Mill $\lambda\left(x_j\widehat{\delta}\right)$. Soit $\gamma = \left(\beta'\ \sigma_\varepsilon\right)$ le vecteur des $K+1$ paramètres à estimer. Le modèle (1.19) s'écrit alors sous la forme :

$$y = \widehat{Z}\gamma + w \tag{1.23}$$

où y désigne le vecteur des N_1 observations de y_i pour lesquelles $y_i > 0$ et où les résidus $w = (w_1...w_{N_1})'$ sont tels que :

$$w_i = v_i + \eta_i = v_i + \sigma_\varepsilon\left[\lambda(x_i\delta) - \lambda\left(x_i\widehat{\delta}\right)\right] \tag{1.24}$$

Le résidu se décompose ainsi en la somme du résidu v_i de la représentation (1.19) et d'un terme provenant de l'erreur d'estimation du paramètre $\delta = \beta/\sigma_\varepsilon$ dans la phase n°1 d'estimation du probit. L'estimateur de Heckman en deux étapes est alors défini par :

$$\widehat{\gamma}_H = \left(\widehat{Z}'\widehat{Z}\right)^{-1}\left(\widehat{Z}'y\right)$$

Amemiya (1983) établit alors lé résultat suivant en ce qui concerne la distribution asymptotique de $\widehat{\gamma}_H$ (cas particulier de Heckman 1979) :

Proposition 1.6. ***L'estimateur en deux étapes de Heckman (1976) est asymptotiquement normalement distribué :***

$$N_1^{-\frac{1}{2}}\left(\widehat{\gamma}_H-\gamma\right)\underset{N\to\infty}{\overset{p}{\longrightarrow}}\mathcal{N}\left(0,V_{\widehat{\gamma}}\right) \tag{1.25}$$

où N_1 ***désigne le nombre d'observations telles que*** $y_i>0$***, avec***

$$V_{\widehat{\gamma}}=\sigma_{\varepsilon}^2\left(Z'Z\right)^{-1}Z'\left[\Sigma+\sigma_{\varepsilon}^2\left(I-\Sigma\right)X\left(X'D_1X\right)^{-1}X'\left(I-\Sigma\right)\right]Z\left(Z'Z\right)^{-1} \tag{1.26}$$

Il faut noter que dans cette expression de $V_{\widehat{\gamma}}$, la seconde matrice dans les crochets provient du fait que le ratio de Mill λ a du être estimé dans une première étape. Si la valeur de ce ratio était connu, la matrice de variance covariance asymptotique deviendrait simplement :

$$V_{\widehat{\gamma}}=\sigma_{\varepsilon}^2\left(Z'Z\right)^{-1}Z'\Sigma Z\left(Z'Z\right)^{-1}$$

Au delà de ces résultats, on vérifie que **l'estimateur de Heckman en deux étapes est asymptotiquement convergent** :

$$\widehat{\gamma}_H\underset{N_1\to\infty}{\overset{p}{\longrightarrow}}\gamma \tag{1.27}$$

On dispose ainsi d'un estimateur convergent et qui ne nécessite dans la première étape que l'utilisation d'un estimateur du maximum de vraisemblance pour un probit simple. Cet estimateur représente donc un gain de capacités de calculs par rapport à l'estimateur du maximum de vraisemblance appliqué directement au modèle Tobit. **Remarquons toutefois, que cet estimateur est biaisé à distance finie en raison de la corrélation entre la perturbation** $w_i=\eta_i+v_i$ **et la variable explicative** $\lambda\left(x_i\widehat{\delta}\right)$.

1.3. Estimation par le Maximum de Vraisemblance

La procédure d'estimation la plus utilisée aujourd'hui est celle du maximum de vraisemblance. En effet, les capacités informatiques sont désormais suffisantes pour envisager l'optimisation des fonctions de vraisemblance associées directement aux modèles Tobit et non plus uniquement aux probit dichotomiques comme dans le cas de la procédure d'Heckman (1976). Commençons par définir la log-vraisemblance associée au modèle Tobit simple :

$$y_i=\begin{cases}y_i^* & \text{si } y_i^*=x_i\beta+\varepsilon_i>0\\ 0 & \text{sinon}\end{cases}\qquad\forall i=1,..N$$

avec $x_i=\left(x_i^1..x_i^K\right)$, $\forall i=1,..,N$, $\beta=\left(\beta_1...\beta_K\right)'\in\mathbb{R}^K$ et où les perturbations ε_i sont distribués selon une loi $N\left(0,\sigma_{\varepsilon}^2\right)$.

1.3.1. Log Vraisemblance dans un modèle Tobit simple

Considérons un échantillon de N observations y_i, noté $y=\left(y_1,..,y_N\right)$. La vraisemblance de ce modèle est définie par :

$$L\left(y,\beta,\sigma_{\varepsilon}^2\right)=\prod_{i:\,y_i=0}\left[1-\Phi\left(\frac{x_i\beta}{\sigma_{\varepsilon}}\right)\right]\prod_{i:\,y_i>0}\left(\frac{1}{\sigma_{\varepsilon}}\right)\phi\left(\frac{y_i-x_i\beta}{\sigma_{\varepsilon}}\right) \tag{1.28}$$

En effet, on sait que si l'on définit une variable dichotomique probit z_i telle que

$$z_i=\begin{cases}1 & \text{si } y_i^*=x_i\beta+\varepsilon_i>0\\ 0 & \text{sinon}\end{cases}\qquad\forall i=1,..N \tag{1.29}$$

alors on peut écrire la probabilité que la variable y_i prenne des valeurs positives sous la forme $Prob\left(z_i=1\right) = Prob\left(\varepsilon_i/\sigma_\varepsilon < x_i\beta/\sigma_\varepsilon\right) = \Phi\left(x_i\beta/\sigma_\varepsilon\right)$. Par conséquent, la probabilité que y_i prenne une valeur nulle s'écrit comme la probabilité complémentaire :

$$Prob\left(y_i=0\right) = Prob\left(z_i=0\right) = 1-\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)$$

Ce qui explique le terme du premier produit de la fonction de vraisemblance (1.28). Le second terme de cette expression correspond tout simplement au produit des lois marginales des variables y_i positives. On sait que si $y_i > 0$, on a part définition $y_i = y_i^* = x_i\beta + \varepsilon_i$ où les perturbations ε_i sont distribués selon une loi $N\left(0,\sigma_\varepsilon^2\right)$. On en déduit que les variables y_i sont distribuées selon une loi normale $N\left(x_i\beta,\sigma_\varepsilon^2\right)$. Ainsi, la loi marginale d'une observation y_i positive est définie par la quantité :

$$\frac{1}{\sigma_\varepsilon\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{y_i - x_i\beta}{\sigma_\varepsilon}\right)^2\right] = \left(\frac{1}{\sigma_\varepsilon}\right)\phi\left(\frac{y_i-x_i\beta}{\sigma_\varepsilon}\right)$$

où $\phi\left(.\right)$ désigne la fonction de densité associée à loi normale centrée réduite.

On en déduit l'écriture de la log-vraisemblance :

Proposition 1.7. ***La log-vraisemblance concentrée associée à un échantillon*** $y = (y_1,..,y_N)$ ***dans un modèle Tobit simple s'écrit :***

$$\log L\left(y,\beta,\sigma_\varepsilon^2\right) = \sum_{i:\,y_i=0}\log\left[1-\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\right] - \frac{N_1}{2}\log\left(\sigma_\varepsilon^2\right) - \frac{1}{2\sigma_\varepsilon^2}\sum_{i:\,y_i>0}\left(y_i-x_i\beta\right)^2 \tag{1.30}$$

où N_1 ***désigne le nombre d'observations pour lesquelles*** $y_i > 0$.

En effet, on sait que la log-vraisemblance est définie par :

$$\begin{aligned}
\log L\left(y,\beta,\sigma_\varepsilon^2\right) &= \sum_{i:\,y_i=0}\log\left[1-\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\right] + \sum_{i:\,y_i>0}\log\left[\left(\frac{1}{\sigma_\varepsilon}\right)\phi\left(\frac{y_i-x_i\beta}{\sigma_\varepsilon}\right)\right] \\
&= \sum_{i:\,y_i=0}\log\left[1-\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\right] - \sum_{i:\,y_i>0}\log\left(\sigma_\varepsilon\right) + \sum_{i:\,y_i>0}\log\left[\phi\left(\frac{y_i-x_i\beta}{\sigma_\varepsilon}\right)\right] \\
&= \sum_{i:\,y_i=0}\log\left[1-\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\right] - N_1\log\left(\sigma_\varepsilon\right) + \sum_{i:\,y_i>0}\log\left[\frac{1}{\sqrt{2\pi}}e^{-\frac{\left(y_i-x_i\beta\right)^2}{2\sigma_\varepsilon^2}}\right] \\
&= \sum_{i:\,y_i=0}\log\left[1-\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\right] - N_1\log\left(\sigma_\varepsilon\right) - \frac{1}{2\sigma_\varepsilon^2}\sum_{i:\,y_i>0}\left(y_i-x_i\beta\right)^2 \\
&\quad -\frac{N_1}{2}\log\left(2\pi\right)
\end{aligned}$$

En omettant les termes constants (*log-vraisemblance concentrée*), il vient :

$$\log L\left(y,\beta,\sigma_\varepsilon^2\right) = \sum_{i:\,y_i=0}\log\left[1-\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\right] - N_1\log\left(\sigma_\varepsilon\right) - \frac{1}{2\sigma_\varepsilon^2}\sum_{i:\,y_i>0}\left(y_i-x_i\beta\right)^2$$

Sachant que $N_1\log\left(\sigma_\varepsilon\right) = \left(N_1/2\right)\log\left(\sigma_\varepsilon^2\right)$, on retrouve l'expression (1.30) de la fonction de log vraisemblance.

On en déduit alors l'expression des dérivées premières par rapport à β et à σ_ε^2 :

Definition 1.8. ***Dans le cas d'un modèle Tobit, le gradient associé à la log-vraisemblance s'écrit sous la forme suivante :***

$$\frac{\partial \log L\left(y,\beta,\sigma_{\varepsilon}^{2}\right)}{\partial \beta}=-\frac{1}{\sigma_{\varepsilon}}\sum_{i:\,y_i=0}\left[\frac{\phi\left(\frac{x_i\beta}{\sigma_{\varepsilon}}\right)x_i'}{1-\Phi\left(\frac{x_i\beta}{\sigma_{\varepsilon}}\right)}\right]+\frac{1}{\sigma_{\varepsilon}^{2}}\sum_{i:\,y_i>0}\left(y_i-x_i\beta\right)x_i' \tag{1.31}$$

$$\frac{\partial \log L\left(y,\beta,\sigma_{\varepsilon}^{2}\right)}{\partial \sigma_{\varepsilon}^{2}}=\frac{1}{2\sigma_{\varepsilon}^{3}}\sum_{i:\,y_i=0}\left[\frac{x_i\beta\,\phi\left(\frac{x_i\beta}{\sigma_{\varepsilon}}\right)}{1-\Phi\left(\frac{x_i\beta}{\sigma_{\varepsilon}}\right)}\right]-\frac{N_1}{2\sigma_{\varepsilon}^{2}}+\frac{1}{2\sigma_{\varepsilon}^{4}}\sum_{i:\,y_i>0}\left(y_i-x_i\beta\right)^2 \tag{1.32}$$

Amemiya (1973) démontre que l'estimateur $\widehat{\gamma}=\left(\widehat{\beta}'\ \widehat{\sigma}_{\varepsilon}\right)'$ du maximum de vraisemblance satisfaisant :

$$\widehat{\gamma}=\underset{\{\gamma\}}{\arg\max}\left[\log L\left(y,\gamma\right)\right]=\underset{\{\beta,\sigma_{\varepsilon}\}}{\arg\max}\left[\log L\left(y,\beta,\sigma_{\varepsilon}^{2}\right)\right] \tag{1.33}$$

est convergent et asymptotiquement distribué selon une loi normale de moyenne nulle et de variance égale à l'inverse de la matrice d'information de Fischer :

$$\sqrt{N}\left(\widehat{\gamma}-\gamma_0\right)\underset{N\rightarrow\infty}{\overset{L}{\longrightarrow}}N\left[0,I\ \left(\gamma_0\right)^{-1}\right] \tag{1.34}$$

avec

$$I\left(\gamma\right)=-E\left[\frac{\partial^2\log\ L\left(y,\gamma\right)}{\partial\gamma\partial\gamma'}\right]_{\gamma=\gamma_0} \tag{1.35}$$

où γ_0 désigne la varie valeur du vecteur de paramètres[3]. Nous allons à présent proposer un changement de paramètre permettant d'obtenir une expression de la log-vraisemblance globalement concave, comme dans le cas des modèles logit et probit dichotomiques.

1.3.2. Re-paramétrisation d'Olsen (1978)

Nous avons montré que les estimateurs du maximum de vraisemblance des paramètres d'un modèle Tobit simple, notées respectivement $\widehat{\beta}$ et $\widehat{\sigma}_{\varepsilon}$, sont solution du programme :

$$\max_{\{\beta,\sigma_{\varepsilon}\}}\ \log L\left(y,\beta,\sigma_{\varepsilon}^{2}\right)$$

et vérifient donc par conséquent les conditions nécessaires suivant, correspondant à l'annulation du vecteur gradient de la log-vraisemblance :

$$\left.\frac{\partial \log L\left(y,\beta,\sigma_{\varepsilon}^{2}\right)}{\partial \beta}\right|_{\beta=\widehat{\beta}}=-\frac{1}{\widehat{\sigma}_{\varepsilon}}\sum_{i:\,y_i=0}\left(\frac{\phi\left(\frac{x_i\widehat{\beta}}{\widehat{\sigma}_{\varepsilon}}\right)}{1-\Phi\left(\frac{x_i\widehat{\beta}}{\widehat{\sigma}_{\varepsilon}}\right)}x_i'\right)+\frac{1}{\widehat{\sigma}_{\varepsilon}^{2}}\sum_{i:\,y_i>0}\left(y_i-x_i\widehat{\beta}\right)x_i'=0$$

$$\left.\frac{\partial \log L\left(y,\beta,\sigma_{\varepsilon}^{2}\right)}{\partial \sigma_{\varepsilon}^{2}}\right|_{\sigma_{\varepsilon}^{2}=\widehat{\sigma}_{\varepsilon}^{2}}=\frac{1}{2\widehat{\sigma}_{\varepsilon}^{3}}\sum_{i:\,y_i=0}\left(\frac{\phi\left(\frac{x_i\widehat{\beta}}{\widehat{\sigma}_{\varepsilon}}\right)}{1-\Phi\left(\frac{x_i\widehat{\beta}}{\widehat{\sigma}_{\varepsilon}}\right)}x_i\widehat{\beta}\right)-\frac{N_1}{2\widehat{\sigma}_{\varepsilon}^{2}}+\frac{1}{2\widehat{\sigma}_{\varepsilon}^{4}}\sum_{i:\,y_i>0}\left(y_i-x_i\widehat{\beta}\right)^2=0$$

Pour déterminer les estimateurs $\widehat{\beta}$ et $\widehat{\sigma}_{\varepsilon}$, il convient donc de résoudre ce système de $K+1$ équations non linéaires. Comme dans le cas des modèles probit et logit, il n'existe pas d'expression analytique des solutions de ce programme. *La résolution d'un tel système ne peut*

[3] La formule de la matrice de variance covariance asympotoique des estimateurs du MVC dans la paramétrisation $(\beta,\sigma_{\varepsilon})$ est donnée dans Amemiya (1973).

donc se faire qu'en utilisant une procédure d'optimisation numérique. Nous avons vu dans le premier chapitre, que généralement on recours alors à des algorithmes d'optimisation fondés notamment sur la méthode du gradient (comme l'algorithme de Newton Raphson par exemple).

Amemiya (1973) a démontré que la fonction de vraisemblance du modèle Tobit paramétrée en β et σ_ε n'est pas globalement concave. Cette propriété est alors particulièrement gênante puisque nous savons que les solutions des algorithmes d'optimisation numérique sont alors extrêmement sensibles au problème du choix des conditions initiales.

S'il existe des extrema locaux de la fonction à optimiser, en l'occurrence ici la fonction de log-vraisemblance, il peut arriver que l'algorithme converge vers ces extrema locaux. En effet, si l'on utilise des conditions initiales dans l'algorithme d'optimisation relativement proches des extrema locaux de la fonction de log-vraisemblance, alors il y a des risques que l'algorithme d'optimisation s'arrête en ces points pour lesquels le gradient est nul, mais qui ne maximisent pas de façon globale la fonction de log-vraisemblance. *On risque alors d'obtenir des estimateurs non convergents des vrais paramètres du modèle Tobit, non pas en raison de mauvaises propriétés de la méthode économétrique utilisée (maximum de vraisemblance), mais simplement en raison de la défaillance de l'algorithme d'optimisation numérique utilisé pour maximiser la log-vraisemblance.* Plusieurs solutions, non exclusives les unes des autres, peuvent être apportées à ce problème :

1. La première solution consiste à modifier les valeurs des conditions initiales de l'algorithmes d'optimisation[4] de sorte à vérifier la robustesse des estimations obtenues à la modification de ces valeurs. Si le changement des valeurs initiales ne conduit à aucune modification des estimations des paramètres, cela tend à montrer que l'algorithme a convergé vers un extremum global. Si en revanche, les estimations sont modifiées, cela prouve que la solution précédente n'était pas un extremum global de la fonction de vraisemblance. Mais se pose alors la question de savoir ce qu'il en est pour les nouvelles estimations obtenues ? Correspondent elles à un extremum global de la fonction de la vraisemblance ?

2. La deuxième solution consiste à vérifier la robustesse des estimations au choix de l'algorithme d'optimisation. Généralement, plusieurs algorithmes sont proposés sous les logiciels usuels : simplex, Newton Raphson, Marquadt etc.. Ces algorithmes, fondées sur des méthodes différentes, n'ont pas la même sensibilité au choix des conditions initiales. Ainsi, si pour différents algorithmes, on obtient des estimations relativement proches, cela tend à prouver que ces estimations correspondent au maximum global de la fonction de log-vraisemblance. Si, en revanche, on obtient des estimations sensiblement différentes pour différents algorithmes ayant convergés, cela tend à montrer que certains de ces algorithmes, pour les conditions initiales posées, ne permettent pas d'identifier le maximum global de la vraisemblance. La question qui se pose est alors de savoir quel algorithme doit être privilégié en fonction du problème posé ?

3. *La troisième solution proposée par Olsen (1978) consiste à reparamétriser la fonction de vraisemblance de sorte à garantir sa concavité globale.* Dès lors, on supprime le problème de la sensibilité des solutions des algorithmes au choix des conditions initiales sur les paramètres puisqu'il n'existe qu'un seul extremum global pour la fonction de log-vraisemblance. *Le choix des conditions initiales et de l'algorithme n'affecte alors que la*

[4] Sous Eviexs, cliquez pour cela sur l'onglet *options* dans la fenêtre d'estimation.

vitesse de convergence des procédure d'optimisation, et ne doit pas théoriquement affecter les résultats.

La solution d'Olsen (1978) est ainsi particulièrement habile puisqu'elle supprime le problème en reformulant la log-vraisemblance du modèle Tobit en des paramètres transformés $\theta = \beta/\sigma_\varepsilon$ et $h = \sigma_\varepsilon^{-1}$ de sorte à obtenir une nouvelle expression de la log-vraisemblance re-paramétrée globalement concave.

Proposition 1.9 (Olsen 1978). ***La log-vraisemblance d'un modèle Tobit re-paramétrée en $\theta = \beta/\sigma_\varepsilon$ et $h = \sigma_\varepsilon^{-1}$ est globalement concave :***

$$\log L(y,\theta,h) = \sum_{i:\, y_i=0} \log\left[1-\Phi(x_i\theta)\right] + N_1 \log(h) - \frac{1}{2}\sum_{i:\, y_i>0}(hy_i - x_i\theta)^2 \tag{1.36}$$

où N_1 désigne le nombre d'observations pour lesquelles $y_i > 0$.

Preuve : la matrice hessienne associée à la log-vraisemblance $\log L(y,\theta,h)$ s'écrit sous la forme suivante :

$$\underset{(K+1,K+1)}{H(\theta,h)} = \begin{pmatrix} \frac{\partial^2 \log L(y,\theta,h)}{\partial\theta\partial\theta'} & \frac{\partial^2 \log L(y,\theta,h)}{\partial\theta\partial h} \\ \frac{\partial^2 \log L(y,\theta,h)}{\partial h\partial\theta'} & \frac{\partial^2 \log L(y,\theta,h)}{\partial h^2} \end{pmatrix} \tag{1.37}$$

Olsen (1978) démontre alors que la matrice hessienne $H(\theta,h)$ est égale à la somme de deux matrices telles que :

$$H(\theta,h) = \Delta + \Gamma = \begin{pmatrix} \Psi(\theta,h) & 0 \\ 0 & -\frac{N_1}{h^2} \end{pmatrix} + \begin{pmatrix} -\sum_{i:\, y_i>0} x_i x_i' & \sum_{i:\, y_i>0} x_i y_i \\ \sum_{i:\, y_i>0} y_i x_i' & \sum_{i:\, y_i>0} y_i^2 \end{pmatrix}$$

où le bloc $\Psi(\theta,h)$ de dimension (K,K) est défini par ::

$$\Psi(\theta,h) = \sum_{i:\, y_i=0} \frac{\phi(x_i\theta)}{1-\Phi(x_i\theta)}\left[x_i\theta - \frac{\phi(x_i\theta)}{1-\Phi(x_i\theta)}\right] x_i x_i'$$

avec $x_i\theta - \phi(x_i\theta)\left[1-\Phi(x_i\theta)\right]^{-1} < 0$. En effet, on sait que la quantité $\phi(z) + z\Phi(z)$ correspond à la primitive de la fonction $\Phi(z)$:

$$\phi(z) + z\Phi(z) = \int_{-\infty}^{z} \Phi(t)\, dt \; > z \quad \forall z \in \mathbb{R}$$

On en déduit que $\forall z \in \mathbb{R}$:

$$\phi(z) > z\left[1-\Phi(z)\right] \Longleftrightarrow z - \phi(z)\left[1-\Phi(z)\right]^{-1} < 0$$

Dès lors, puisque $x_i'x_i$ est une matrice définie positive, les deux matrices Δ et Γ sont des matrices définies négatives (cf annexe A.1) : dès lors, la matrice hessienne est égale à la somme de deux matrices définies négatives, elle est donc définie négative. La fonction de log-vraisemblance est donc globalement concave.

Lorsque la log-vraisemblance est paramétrée en h et θ, le gradient s'écrit sous la forme suivante :

Definition 1.10. ***Le gradient associé à la log-vraisemblance d'un modèle Tobit re-paramétrée en $\theta = \beta/\sigma_\varepsilon$ et $h = \sigma_\varepsilon^{-1}$ est :***

$$\frac{\partial \log L(y, \theta, h)}{\partial \theta} = \sum_{i:\, y_i = 0} \frac{\phi(x_i \theta)}{1 - \Phi(x_i \theta)} x_i' - \frac{1}{2} \sum_{i:\, y_i > 0} (h y_i - x_i \theta) x_i' \tag{1.38}$$

$$\frac{\partial \log L(y, \theta, h)}{\partial h} = \frac{N_1}{h} - \frac{1}{2} \sum_{i:\, y_i > 0} (h y_i - x_i \theta) y_i \tag{1.39}$$

Compte tenu du résultat d'Olsen, il est possible en utilisant des algorithmes d'optimisation usuels de déterminer les estimateurs du maximum de vraisemblance des paramètres transformés $\widehat{\theta}$ et $\widehat{h}$. Ces estimateurs sont solutions du programme suivant :

$$\left(\widehat{\theta}' \ \widehat{h}\right) = \max_{\{\theta, h\}} \log L(y, \theta, h)$$

et vérifient naturellement les conditions nécessaires suivantes :

$$\left.\frac{\partial \log L(y, \theta, h)}{\partial \theta}\right|_{\theta = \widehat{\theta}} = \left.\frac{\partial \log L(y, \theta, h)}{\partial h}\right|_{h = \widehat{h}} = 0$$

On en déduit alors les estimateurs des paramètres du modèles Tobit originel puisque l'on a $\widehat{\theta} = \widehat{\beta}/\widehat{\sigma}_\varepsilon$ et $\widehat{h} = \widehat{\sigma}_\varepsilon^{-1}$:

$$\widehat{\sigma}_\varepsilon = \widehat{h} \qquad \widehat{\beta} = \widehat{\theta}\ \widehat{\sigma}_\varepsilon \tag{1.40}$$

La matrice de variance covariance asymptotique des estimateurs $\widehat{\sigma}_\varepsilon$ et $\widehat{\beta}$ se déduit alors de celle de $\widehat{\theta}$ et $\widehat{\beta}$, qui s'exprime en fonction de la matrice hessienne $H\left(\widehat{\theta}, \widehat{h}\right)$ selon les formules usuelles.

1.4. Application

Considérons tout d'abord une application sur données simulées qui nous permettra par la suite d'évaluer la portée des biais. On simule un échantillon de 1000 points satisfaisant les propriétés suivantes :

$$y_i = \begin{cases} y_i^* & \text{si } y_i^* = \alpha + \beta x_i + \varepsilon_i > 0 \\ 0 & \text{sinon} \end{cases} \qquad \forall i = 1, ..N$$

avec $x_i \in \mathbb{R}$, $\forall i = 1, .., N$, pour une taille d'échantillon $N = 1000$ et où les perturbations ε_i sont distribués selon une loi $N\left(0, \sigma_\varepsilon^2\right)$. On pose la valeur suivante des paramètres :

$$\alpha = 1 \quad \beta = 0.8 \quad \sigma_\varepsilon^2 = 1$$

On suppose ici que la variable explicative x_i satisfait l'hypothèse de Goldberger (1981) : la variable explicative x_i est distribuée selon une loi normale $N(0, \Omega)$, avec $\Omega = 1$, et est indépendante du résidu, $cov(\varepsilon_i x_i) = 0$. Le programme permettant de simuler la série observable y_i est fourni en annexe (A.2). Commençons par estimer les paramètres α, β et σ_ε^2 par une méthode de maximisation de la vraisemblance standard. Les résultats sont représentés dans la figure (1.2) :

Eviews indique tout d'abord que l'échantillon simulé comporte 781 observations pour lesquelles $y_i > 0$ et 219 observations censurées à gauche, c'est à dire pour lesquels $y_i = 0$. On vérifie tout d'abord que l'algorithme d'optimisation numérique de la maximisation de la vraisemblance a convergé après 5 itérations.. Compte tenu de la taille d'échantillon N relativement importante,

Figure 1.2: Estimation Modèle Tobit Simple par Maximum de Vraisemblance

Dependent Variable: Y
Method: ML - Censored Normal (TOBIT)
Date: 11/15/02 Time: 16:20
Sample: 1 1000
Included observations: 1000
Left censoring (value) at zero
Convergence achieved after 5 iterations
QML (Huber/White) standard errors & covariance

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.965009	0.496771	1.942562	0.0521
X	0.793976	0.348861	2.275910	0.0229
		Error Distribution		
SCALE:C(3)	0.970557	0.170952	5.677378	0.0000

R-squared	0.398176	Mean dependent var	1.145478
Adjusted R-squared	0.396969	S.D. dependent var	1.039315
S.E. of regression	0.807081	Akaike info criterion	2.526896
Sum squared resid	649.4257	Schwarz criterion	2.541620
Log likelihood	-1260.448	Hannan-Quinn criter.	2.532492
Avg. log likelihood	-1.260448		
Left censored obs	219	Right censored obs	0
Uncensored obs	781	Total obs	1000

les réalisations des estimateurs $\widehat{\alpha} = 0.965$ et $\widehat{\beta} = 0.793$ sont très proches des vraies valeurs $\alpha = 1$ et $\beta = 0.8$. Eviews fournit en outre un estimation de la variance résiduelle, comme tenu de la distribution choisie (en l'occurrence une loi normale dans le cas d'un modèle Tobit simple) : la réalisation de l'estimateur $\widehat{\sigma}_{\varepsilon}^2$ est alors égale à 0.97, valeur relativement proche de la vraie valeur de la variance $\sigma_{\varepsilon}^2 = 1$. Les z statistiques correspondant aux tests de nullité des paramètres nous permettent de rejeter l'hypothèse nulle au seuil de 5% pour les trois paramètres α, β et σ_{ε}^2.

Comparons la réalisation de ces estimateurs du maximum de vraisemblance à celles obtenues par les estimateurs des MCO appliqués à l'échantillon complet, notés $\widehat{\alpha}_{LS}$, $\widehat{\beta}_{LS}$ et $\widehat{\sigma}_{\varepsilon,LS}^2$ reportés sur la figure (1.3).

On vérifie que l'estimation par les MCO sur les 1000 points des paramètres α et β donne des résultats largement moins bons que ceux obtenus par maximum de vraisemblance, puisque nous avons vu précédemment que ces estimateurs sont biaisés. En effet pour une vraie valeur $\beta = 0.8$, la réalisation de l'estimateur des MCO est, dans notre expérience, de 0.6253.

Nous avions vu que sous l'hypothèse de normalité des variables x_i (hypothèse de Goldberger 1981), l'estimateur des MCO du paramètre β vérifie :

$$\widehat{\beta}_{LS} \underset{N\to\infty}{\overset{p}{\longrightarrow}} \beta \times \Phi\left(\frac{\alpha}{\sigma_y}\right) \tag{1.41}$$

Dans le cas de notre expérience, sachant que $\alpha = 1$ et que :

$$\sigma_y^2 = \sigma_{\varepsilon}^2 + \beta'\Omega\beta = \sigma_{\varepsilon}^2 + \beta^2\Omega = 1 + 0.8^2 \times 1 = 1.64$$

Figure 1.3: Estimation par les MCO sur l'échantillon complet

Dependent Variable: Y
Method: Least Squares
Date: 11/15/02 Time: 15:48
Sample: 1 1000
Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.127753	0.025844	43.63672	0.0000
X	0.625391	0.025140	24.87645	0.0000

R-squared	0.382746	Mean dependent var	1.145478
Adjusted R-squared	0.382127	S.D. dependent var	1.039315
S.E. of regression	0.816953	Akaike info criterion	2.435527
Sum squared resid	666.0769	Schwarz criterion	2.445342
Log likelihood	-1215.763	F-statistic	618.8378
Durbin-Watson stat	1.999318	Prob(F-statistic)	0.000000

on en déduit que

$$\widehat{\beta}_{LS} \underset{N\rightarrow\infty}{\overset{p}{\longrightarrow}} 0.8\times\Phi\left(\frac{1}{\sqrt{1.64}}\right)=0.8\times\Phi\left(0.78087\right)=0.6260$$

Ainsi, on sait que théoriquement l'estimateur $\widehat{\beta}_{LS}$ converge en probabilité vers la valeur 0.6260. On vérifie en effet sur la figure (1.3), pour une taille d'échantillon $N=1000$ relativement importante, que la réalisation de $\widehat{\beta}_{LS}=0.6253$ est très proche de cette valeur asymptotique.

Nous avions vu en outre, toujours sous l'hypothèse de normalité des variables explicatives x_i, que l'estimateur des MCO corrigé $\widehat{\beta}_{LS}^{c}=(N/N_1)\times\widehat{\beta}_{LS}$ est convergent :

$$\widehat{\beta}_{LS}^{c}=\frac{N}{N_1}\times\widehat{\beta}_{LS} \underset{N\rightarrow\infty}{\overset{p}{\longrightarrow}} \beta$$

Dans le cas de notre simulation, la réalisation de cette estimateur vaut :

$$\widehat{\beta}_{LS}^{c}=\frac{N}{N_1}\times\widehat{\beta}_{LS}=\frac{1000}{781}\times 0.6253=.80064$$

Cette réalisation est en effet très proche de la vraie valeur $\beta=0.8$. On remarque que pour notre échantillon simulé, la réalisation de l'estimateur des MCO corrigé est plus proche de la vraie valeur que l'estimateur du MV.

****

**** 1°) Faire estimation MCO sur partie positive de la distribution

**** 2°) Introduire N simulations sur $\widehat{\beta}_{LS}$, $\widehat{\beta}_{LS}^{c}$, $\widehat{\beta}_{y>0}$, $\widehat{\beta}_{MV}$ et $\widehat{\beta}_{Hec}$ en contrôlant le pourcentage de données censurées : Matlab

**** 3°) Répartir les applications Eviews ou Limdep sur les différentes sections ?

****

1.5. Effets marginaux

Supposons que l'on dispose d'un estimateur convergent $\widehat{\beta}$ des paramètres β et d'un estimateur convergent $\widehat{\sigma}_{\varepsilon}^{2}$ de la variance des résidus. On cherche à mesurer les effets marginaux.

Definition 1.11. ***Les effets marginaux dans un modèle de régression censuré correspondent à la déformation des prévisions sur une variable continue engendrée par une variation d'une unité d'une des variables explicatives.***

Il y alors plusieurs prévisions possible dans le cas du modèle Tobit suivant que l'on s'intéresse à la variable censurée y_i ou à la variable latente y_i^*. En effet, trois cas peuvent apparaître :

1. Soit l'on considère la prévision sur **la variable latente** représentée par l'espérance conditionnelle $\forall i = 1, .., N$:
$$E\left(y_i^* / x_i\right) = x_i \beta \tag{1.42}$$
2. Soit l'on considère la prévision sur **la variable dépendante** représentée par l'espérance conditionnelle $\forall i = 1, .., N$:
$$E\left(y_i / x_i\right) = \Phi\left(\frac{x_i \beta}{\sigma_\varepsilon}\right) x_i \beta + \sigma_\varepsilon \phi\left(\frac{x_i \beta}{\sigma_\varepsilon}\right) \tag{1.43}$$
3. Soit l'on considère la prévision sur **la variable dépendante censurée** représentée par l'espérance conditionnelle $\forall i = 1, .., N$:
$$E\left(y_i / x_i, y_i > 0\right) = E\left(y_i^* / x_i, , y_i^* > 0\right) = x_i \beta + \sigma_\varepsilon \lambda\left(\frac{x_i \beta}{\sigma_\varepsilon}\right) \tag{1.44}$$

On peut ainsi déterminer différents effets marginaux suivant que l'on considère l'une ou l'autre de ces prévisions. Tout d'abord si l'on considère la prévision sur la variable latente, on obtient tout simplement un effet marginal mesuré par la dérivé partielle de l'espérance conditionnelle $E\left(y_i^* / x_i\right)$ par rapport à une composante quelconque du vecteur des variables explicatives x_i.

Definition 1.12. ***L'effet marginal d'une variation unitaire de la*** $k^{ème}$ ***variable explicative*** $x_i^{(k)}$, $\forall k = 1, .., K$, ***sur la prévision de la variable latente*** y_i^* ***est mesuré par la quantité :***
$$\frac{\partial E\left(y_i^* / x_i\right)}{\partial x_i^{(k)}} = \beta^{(k)} \qquad \forall i = 1, .., N \tag{1.45}$$
ou par l'élasticité $\varepsilon_{y_i^* / x_i^{[k]}}$ ***:***
$$\varepsilon_{y_i^* / x_i^{[k]}} = \frac{\partial E\left(y_i^* / x_i\right)}{\partial x_i^{(k)}} \frac{x_i^{(k)}}{E\left(y_i^* / x_i\right)} = \frac{x_i^{(k)} \beta^{(k)}}{x_i \beta} \qquad \forall i = 1, .., N \tag{1.46}$$

Ainsi, une variation de 1% de la $k^{ème}$ variable explicative $x_i^{(k)}$ pour le $i^{ème}$ individu, modifie la prévision de la variable latente y_i^* pour ce même individu de $\varepsilon_{y_i^* / x_i^{[k]}}$ pour cent. On peut alors calculer une élasticité moyenne $\overline{\varepsilon}_{y_i^* / x_i^{[k]}}$ sur l'ensemble des N individus telle que :
$$\overline{\varepsilon}_{y_i^* / x_i^{[k]}} = \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{y_i^* / x_i^{[k]}} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{x_i^{(k)} \beta^{(k)}}{x_i \beta}\right)$$

Considérons à présent la prévision sur la variable dépendante non censurée. De la même façon, l'effet marginal est mesuré par la dérivé partielle de l'espérance conditionnelle $E\left(y_i / x_i\right)$ par rapport à une composante quelconque du vecteur des variables explicatives x_i.

Definition 1.13. ***L'effet marginal d'une variation unitaire de la*** $k^{\grave{e}me}$ ***variable explicative*** $x_i^{(k)}$, $\forall k = 1,..,K$, ***sur la prévision de la variable dépendante*** y_i ***est mesuré par la quantité :***

$$\frac{\partial E\left(y_i / x_i\right)}{\partial x_i^{(k)}} = \Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\beta^{(k)} \qquad \forall i = 1,..,N \tag{1.47}$$

ou par l'élasticité $\varepsilon_{y_i/x_i^{[k]}}$ ***:***

$$\varepsilon_{y_i/x_i^{[k]}} = \frac{\partial E\left(y_i / x_i\right)}{\partial x_i^{(k)}}\frac{x_i^{(k)}}{E\left(y_i / x_i\right)} = \frac{x_i^{(k)}\beta^{(k)}}{x_i\beta + \sigma_\varepsilon\lambda\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)} \qquad \forall i = 1,..,N \tag{1.48}$$

Preuve : A partir de l'espérance conditionnelle $E\left(y_i / x_i\right)$ déterminons l'effet marginal associé à $x_i^{(k)}$.

$$\begin{aligned}
\frac{\partial E\left(y_i / x_i\right)}{\partial x_i^{(k)}} &= \frac{\partial}{\partial x_i^{(k)}}\left[\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)x_i\beta + \sigma_\varepsilon\phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\right] \\
&= \frac{\partial\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)}{\partial x_i^{(k)}}\frac{\beta^{(k)}}{\sigma_\varepsilon}x_i\beta + \Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\beta^{(k)} + \sigma_\varepsilon\frac{\partial\phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)}{\partial x_i^{(k)}}\frac{\beta^{(k)}}{\sigma_\varepsilon} \\
&= \beta^{(k)}\left[\frac{\partial\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)}{\partial x_i^{(k)}}\frac{x_i\beta}{\sigma_\varepsilon} + \Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right) + \frac{\partial\phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)}{\partial x_i^{(k)}}\right]
\end{aligned}$$

Si l'on pose $z = x_i\beta/\sigma_\varepsilon$, on obtient :

$$\frac{\partial E\left(y_i / x_i\right)}{\partial x_i^{(k)}} = \beta^{(k)}\left[\frac{\partial\Phi\left(z\right)}{\partial x_i^{(k)}}z + \frac{\partial\phi\left(z\right)}{\partial x_i^{(k)}} + \Phi\left(z\right)\right]$$

Or, on sait que la quantité $\phi\left(z\right) + z\Phi\left(z\right)$ correspond à la primitive de la fonction $\Phi\left(z\right)$:

$$\phi\left(z\right) + z\Phi\left(z\right) = \int_{-\infty}^{z}\Phi\left(t\right)dt \qquad \forall z \in \mathbb{R}$$

Dès lors, par dérivation par rapport à une composante $z^{(k)}$ on obtient :

$$\begin{aligned}
&\frac{\partial\phi\left(z\right)}{\partial z^{(k)}} + \frac{\partial\left[z\Phi\left(z\right)\right]}{\partial z^{(k)}} = \Phi\left(z\right) \\
\Longleftrightarrow\ &\frac{\partial\phi\left(z\right)}{\partial z^{(k)}} + \frac{\partial z}{\partial z^{(k)}}\Phi\left(z\right) + z\frac{\partial\Phi\left(z\right)}{\partial z^{(k)}} = \frac{\partial z}{\partial z^{(k)}}\Phi\left(z\right) \\
\Longleftrightarrow\ &\frac{\partial\phi\left(z\right)}{\partial z^{(k)}} + z\frac{\partial\Phi\left(z\right)}{\partial z^{(k)}} = 0
\end{aligned}$$

Ainsi, on obtient finalement que :

$$\frac{\partial E\left(y_i / x_i\right)}{\partial x_i^{(k)}} = \beta^{(k)}\Phi\left(z\right) = \beta^{(k)}\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)$$

En ce qui concerne l'élasticité $\varepsilon_{y_i/x_i^{[k]}}$ on montre que celle-ci est définie par la quantité suivante :

$$\begin{aligned}
\varepsilon_{y_i/x_i^{[k]}} &= \Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\beta^{(k)}\frac{x_i^{(k)}}{\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)x_i\beta + \sigma_\varepsilon\phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)} \\
&= \frac{1}{x_i\beta + \sigma_\varepsilon\lambda\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)}x_i^{(k)}\beta^{(k)}
\end{aligned}$$

Ainsi, une variation de 1% de la $k^{ème}$ variable explicative $x_i^{(k)}$ pour le $i^{ème}$ individu, modifie la prévision de la variable dépendante y_i pour ce même individu de $\varepsilon_{y_i/x_i^{[k]}}$ pour cent. On peut alors calculer une élasticité moyenne $\overline{\varepsilon}_{y_i/x_i^{[k]}}$ sur l'ensemble des N individus telle que :

$$\overline{\varepsilon}_{y_i/x_i^{[k]}} = \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{y_i/x_i^{[k]}} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i^{(k)}\beta^{(k)}}{x_i\beta+\sigma_\varepsilon\lambda\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)}\right)$$

De façon générale, on montre que $\left|\overline{\varepsilon}_{y_i^*/x_i^{[k]}}\right| > \left|\overline{\varepsilon}_{y_i/x_i^{[k]}}\right|$.

McDonald et Moffit (1980) ont proposé une décomposition particulièrement intéressante de l'effet marginal associé à la prévision sur la variable dépendante y_i. Cette décomposition est la suivante :

$$\begin{aligned}\frac{\partial E\left(y_i/\,x_i\right)}{\partial x_i^{(k)}} &= \Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\beta^{(k)}\left\{1-\lambda_i\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\left[\frac{x_i\beta}{\sigma_\varepsilon}+\lambda_i\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\right]\right\}\\ &\quad+\beta^{(k)}\phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\left[\frac{x_i\beta}{\sigma_\varepsilon}+\lambda_i\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\right]\end{aligned}$$

Dès lors, l'effet marginal d'une variation unitaire de la $k^{ème}$ variable explicative $x_i^{(k)}$, $\forall k = 1,..,K$, sur la prévision de la variable dépendante y_i peut se décomposer comme la somme de deux éléments :

Remark 3. ***La variation de $x_i^{(k)}$ a deux effets sur la prévision de la variable dépendante y_i représentés par la décomposition de McDonald et Moffit (1980):***

$$\frac{\partial E\left(y_i/\,x_i\right)}{\partial x_i^{(k)}} = Prob\left(y_i>0\right)\frac{\partial E\left(y_i/\,x_i,y_{i>0}\right)}{\partial x_i^{(k)}} + E\left(y_i/\,x_i,y_{i>0}\right)\frac{\partial Prob\left(y_i>0\right)}{\partial x_i^{(k)}}$$

1. ***D'une part, la variation de $x_i^{(k)}$ modifie l'espérance conditionnelle de y_i dans la partie positive de la distribution.***
2. ***D'autre part, la variation de $x_i^{(k)}$ affecte la probabilité que l'observation y_i appartienne à cette partie de la distribution.***

Au passage, cette décomposition nous donne la $3^{ème}$ mesure de l'effet marginal : celle relative à la prévision de la variable dépendante sur la partie positive de la distribution :

$$\frac{\partial E\left(y_i/\,x_i,y_{i>0}\right)}{\partial x_i^{(k)}} = \beta^{(k)}\left\{1-\lambda_i\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\left[\frac{x_i\beta}{\sigma_\varepsilon}+\lambda_i\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\right]\right\}$$

où $\lambda\left(x\right)=\phi\left(x\right)/\Phi\left(x\right)$ désigne le ratio de Mill.

****

Application : construire les différentes EM et commentez (exemple éco)

Utiliser la simulation ou Utiliser exemple Eco :

1. calculer EM1 et élasticité sur y^*
2. calculer EM2 et élasticité sur y
3. Décomposer EM2 par McDonald et Moffit

****

1.6. Propriétés de l'estimateur du MV sous des hypothèses non standard

Nous allons à présent nous intéresser aux propriétés de l'estimateur du MV sous principales hypothèses en présentant chaque fois les tests appropriés :

1. Hypothèse d'hétéroscédasticité
2. Hypothèse de non normalité

Commençons par évoquer les problèmes d'hétéroscédasticité.

1.6.1. Hétéroscédasticité

De nombreuses études ont été consacrées au problème de l'hétéroscédasticité dans le cadre des modèles Tobit simple. Hurd (1979) a ainsi évalué les biais asymptotiques de l'estimateur du MV d'un **modèle Tobit simple tronqué** en présence de différentes formes d'hétéroscédasticité. Rappelons que dans le cas d'un modèle **Tobit simple tronqué** , on ne dispose que d'observations pour les individus pour lesquels $y_i^* > 0$. La vraisemblance s'écrit alors sous la forme :

$$L\left(y,\beta,\sigma_\varepsilon^2\right) = \prod_{i=1}^N \Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)^{-1}\left(\frac{1}{\sigma_\varepsilon}\right)\phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right) \tag{1.49}$$

Hurd considère une certaine forme d'hétéroscédasticité en générant deux sous échantillons : un échantillon de taille rN, avec $r \in [0,1]$, d'observations pour lesquelles $\sigma_\varepsilon^2 = \sigma_1^2$ et un second échantillon de taille $(1-r)\,N$ d'observations pour lesquelles $\sigma_\varepsilon^2 = \sigma_2^2 \neq \sigma_1^2$. Il étudie alors la déformation de la limite en probabilité de l'estimateur du MV, noté $plim\,\widehat{\beta}$, en fonction des valeurs de σ_1^2 en considérant $\sigma_2^2 = 1$ et $r = 0.5$. Hurd démontre ainsi l'existence de biais asymptotiques sur l'estimateur du MV en présence d'hétéroscédasticité et il constate que ces biais peuvent être très importants pour certaines valeurs de σ_1^2. Reprenant la même approche, **Arabmazar et Schmidt (1981) montre que les biais asymptotiques de l'estimateur du MV sont beaucoup moins important dans le cadre d'un modèle Tobit simple censuré**, tel que celui que l'on a vu jusqu'à présent. Les résultats de ces deux études illustrent parfaitement le sens général des résultats de cette littérature.

Proposition 1.14. ***De façon générale, on montre que l'estimateur du MV en présence d'hétéroscédasticité est asymptotiquement biaisé. L'importance des biais asymptotiques croît avec le degré de censure des données.***

Il est toutefois difficile d'aboutir à une conclusion plus précise que cette proposition dans la mesure où les différentes études proposées sur ce thème diffèrent très sensiblement sur la représentation retenue de l'hétéroscédasticité. Les modèles utilisés sont en effet très spécifique. Il faut ainsi simplement retenir que l'hétéroscédasticité pose un sérieux problème d'estimation des modèles Tobit simples.

La question qui se pose est alors de savoir comment tester l'hétéroscédasticité ? Considérons tout d'abord une forme particulière d'hétéroscédasticité.

Hypothèses *Soit un modèle Tobit hétéroscédastique tel que $\forall i = 1,..,N$:*

$$\sigma_{\varepsilon,i}^2 = \sigma_{\varepsilon,i}^2\left(\alpha\right) = \sigma_\varepsilon^2 \exp\left(w_i\alpha\right) \tag{1.50}$$

où $\alpha = (\alpha_1 .. \, \alpha_P)'$ et où $w_i = \left(w_i^{(1)}...w_i^{(P)}\right) \in \mathbb{R}^P$ est un vecteur de caractéristiques.

Cette spécification est suffisamment générale pour englober différentes configurations d'hétéroscédasticité. En particulier, lorsque $\alpha = 0$, on retrouve le modèle Tobit simple homoscédastique. Sous cette hypothèse, la log-vraisemblance concentrée du modèle Tobit simple s'écrit alors :

$$\begin{aligned}\log L\left(y,\beta,\sigma_{\varepsilon}^{2},\alpha\right) &= \sum_{i:\,y_i=0}\log\left[1-\Phi\left(\frac{x_i\beta}{\sigma_{\varepsilon,i}\left(\alpha\right)}\right)\right]\\ &\quad -\frac{N_1}{2}\log\left[\sigma_{\varepsilon,i}^{2}\left(\alpha\right)\right]-\frac{1}{2\sigma_{\varepsilon,i}^{2}\left(\alpha\right)}\sum_{i:\,y_i>0}\left(y_i-x_i\beta\right)^2\end{aligned}$$

Les estimateurs du MV des paramètres du Tobit hétéroscédastique, notés $\widehat{\beta},\widehat{\sigma}_{\varepsilon}^{2}$ et $\widehat{\alpha}$ vérifient alors respectivement les conditions suivantes :

$$\left.\frac{\partial\log L\left(y,\beta,\sigma_{\varepsilon}^{2},\alpha\right)}{\partial\gamma}\right|_{\gamma=\widehat{\gamma}}=0\qquad\forall\gamma=\left\{\beta,\sigma_{\varepsilon}^{2},\alpha\right\}$$

où les composantes du gradient de la fonction de log-vraisemblance en β et σ_{ε}^{2} correspondent à celles définies pour le modèle Tobit simple (cf. proposition 1.8) lorsque $\alpha = 0$, c'est à dire dans le cas homoscédastique.

$$\left.\frac{\partial\log L\left(y,\beta,\sigma_{\varepsilon}^{2},\alpha\right)}{\partial\beta}\right|_{\alpha=0}=-\frac{1}{\sigma_{\varepsilon}}\sum_{i:\,y_i=0}\left[\frac{\phi\left(\frac{x_i\beta}{\sigma_{\varepsilon}}\right)x_i'}{1-\Phi\left(\frac{x_i\beta}{\sigma_{\varepsilon}}\right)}\right]+\frac{1}{\sigma_{\varepsilon}^{2}}\sum_{i:\,y_i>0}\left(y_i-x_i\beta\right)x_i'=\sum_{i=1}^{N}a_ix_i'$$

$$\left.\frac{\partial\log L\left(y,\beta,\sigma_{\varepsilon}^{2},\alpha\right)}{\partial\sigma_{\varepsilon}^{2}}\right|_{\alpha=0}=\frac{1}{2\sigma_{\varepsilon}^{3}}\sum_{i:\,y_i=0}\left[\frac{x_i\beta\,\phi\left(\frac{x_i\beta}{\sigma_{\varepsilon}}\right)}{1-\Phi\left(\frac{x_i\beta}{\sigma_{\varepsilon}}\right)}\right]-\frac{N_1}{2\sigma_{\varepsilon}^{2}}+\frac{1}{2\sigma_{\varepsilon}^{4}}\sum_{i:\,y_i>0}\left(y_i-x_i\beta\right)^2=\sum_{i=1}^{N}b_i$$

Sous ces hypothèses, un test naturel de l'hypothèse d'hétéroscédasticité consiste donc à tester la nullité du vecteur α, puisque si $\alpha = 0$, on a $\sigma_{\varepsilon,i}^{2}=\sigma_{\varepsilon}^{2}\ \forall i=1,..,N$. *Ainsi sous les hypothèses précédentes, le test de l'hypothèse nulle d'homoscédasticité revient au test bilatéral suivant :*

$$H_0\ :\ \alpha=0 \tag{1.51}$$

$$H_a\ :\ \alpha\neq0 \tag{1.52}$$

Plusieurs méthodes sont envisageable pour mener à bien ce test sur les paramètres. Greene (1997) propose d'utiliser un test du multiplicateur de Lagrange[5] (cf. chapitre 1).

Definition 1.15. ***La statistique*** LM ***du test de l'hypothèse nulle d'homoscédasticité*** $H_0:\alpha=0$ ***est définie par :***

$$LM=\left(\left.\frac{\partial\log L\left(y,\widehat{\beta},\widehat{\sigma}_{\varepsilon},\alpha\right)}{\partial\alpha}\right|_{\alpha=0}\right)'Q_{\alpha'\alpha}\left(\left.\frac{\partial\log L\left(y,\widehat{\beta},\widehat{\sigma}_{\varepsilon},\alpha\right)}{\partial\alpha}\right|_{\alpha=0}\right) \tag{1.54}$$

[5] La statistique LM du multiplicateur de Lagrange associée au test unidirectionnel $H_0:\gamma=a\in\mathbb{R}^k$ contre $H_1:\gamma\neq a$ admet la loi suivante sous H_0 :

$$LM=\left(\left.\frac{\partial\log L\left(y,\gamma\right)}{\partial\gamma'}\right|_{\gamma=\widehat{\gamma}^c}\right)'\widehat{I}^{-1}\left(\left.\frac{\partial\log L\left(y,\gamma\right)}{\partial\gamma'}\right|_{\gamma=\widehat{\gamma}^c}\right)\underset{N\rightarrow\infty}{\overset{L}{\longrightarrow}}\chi^2\left(k\right) \tag{1.53}$$

où $\widehat{\gamma}$et $\widehat{\gamma}^c$désignent respectivement les estimateurs non contraint et contraint de γ.

où $\widehat{\beta}$ et $\widehat{\sigma}_{\varepsilon}$ désignent les estimateurs du MV des paramètres β et σ_{ε}^2 obtenus sous l'hypothèse nulle $\alpha = 0$, et où la matrice $Q_{\alpha'\alpha}$ désigne le bloc de dimension (P, P) correspondant au vecteur de paramètre α de la matrice inverse de la matrice d'information de Fischer estimée sous H_0 :

$$\underset{(K+P+1,K+P+1)}{I\left(\widehat{\beta}, \widehat{\sigma}_{\varepsilon}^2, \alpha\right)}^{-1}\Bigg|_{\alpha=0} = \left(\begin{array}{c|c|c} \underset{(K,K)}{Q_{\beta\beta'}} & \underset{(K,1)}{Q_{\beta\sigma_{\varepsilon}^2}} & \underset{(K,P)}{Q_{\beta'\alpha}} \\ \hline \underset{(1,K)}{Q_{\beta'\sigma_{\varepsilon}^2}} & \underset{(1,1)}{Q_{\sigma_{\varepsilon}^2}} & \underset{(1,P)}{Q_{\alpha'\sigma_{\varepsilon}^2}} \\ \hline \underset{(P,K)}{Q_{\alpha'\beta}} & \underset{(P,1)}{Q_{\alpha\sigma_{\varepsilon}^2}} & \underset{(P,P)}{Q_{\alpha\alpha'}} \end{array}\right) \tag{1.55}$$

On montre alors que sous H_0 cette statistique converge en loi :

$$LM \xrightarrow[N\to\infty]{L} \chi^2(P) \tag{1.56}$$

où P rappelons-le désigne la dimension du vecteur de varaibles explicatives w_i expliquant la variance indivudelle $\sigma_{\varepsilon,i}^2(\alpha)$. Ainsi, si la réalisation de la statistique LM est supérieure au fractile de la loi du chi-2 à P degrés de liberté, alors on rejette l'hypothèse nulle d'homoscédasticité. Les résidus du modèle Tobit sont hétéroscédastiques : les estimateurs du MV des paramètres β et σ_{ε}^2 sont asymptotiquement biaisés selon les résultats d'Arabmazar et Schmidt (1981).

Quelle que soit la nature du modèle, il existe une autre façon de construire la statistique LM. Celle-ci peut en effet s'écrire en fonction de la matrice $G\left(\beta, \sigma_{\varepsilon}^2, \alpha\right)$ de dimension $(N, K + P + 1)$ contenant les dérivés de la log-vraisemblance évaluées *pour chaque observation sous l'hypothèse* H_0.Soit $G\left(\beta, \sigma_{\varepsilon}^2, \alpha\right)$ le vecteur du gradient *évalué sous* H_0 :

$$\underset{(N,K+1+P)}{G\left(\beta, \sigma_{\varepsilon}^2, \alpha\right)}\Bigg|_{\alpha=0} = \begin{pmatrix} \underset{(1,K+P+1)}{g_1\left(\beta, \sigma_{\varepsilon}^2, \alpha\right)} \\ \ldots \\ \underset{(1,K+P+1)}{g_N\left(\beta, \sigma_{\varepsilon}^2, \alpha\right)} \end{pmatrix} \tag{1.57}$$

où les vecteurs g_i correspondent au gradient de la fonction de la log-vraisemblance évalués sous l'hypothèse nulle $\alpha = 0$ et pour chaque observation individuelle y_i :

$$\underset{(1,K+P+1)}{g_i\left(\beta, \sigma_{\varepsilon}^2, \alpha\right)} = \left(\left.\frac{\partial \log L\left(y_i, \beta, \sigma_{\varepsilon}^2, \alpha\right)}{\partial \beta}\right|_{\alpha=0} \quad \left.\frac{\partial \log L\left(y_i, \beta, \sigma_{\varepsilon}^2, \alpha\right)}{\partial \sigma_{\varepsilon}^2}\right|_{\alpha=0} \quad \left.\frac{\partial \log L\left(y_i, \beta, \sigma_{\varepsilon}^2, \alpha\right)}{\partial \alpha}\right|_{\alpha=0} \right)$$

Greene (1997) montre que les vecteurs $g_i(.)$ s'écrivent sous la forme :

$$\underset{(1,K+P+1)}{g_i\left(\beta, \sigma_{\varepsilon}^2, \alpha\right)} = \left(\underset{(1,K)}{a_i x_i'} \quad \underset{(1,1)}{b_i} \quad \underset{(1,P)}{\sigma_{\varepsilon}^2 b_i w_i'} \right) \tag{1.58}$$

où les scalaires a_i et b_i sont définies par les composantes du gradient associé à la log-vraisemblance du modèle Tobit simple (cf. proposition 1.8) d'une observation donnée y_i, $\forall i = 1, ..N$.

$$a_i = -\frac{1}{\sigma_{\varepsilon}}(1 - z_i)\,\widetilde{\lambda}\left(\frac{x_i\beta}{\sigma_{\varepsilon}}\right) + \frac{z_i}{\sigma_{\varepsilon}}(y_i - x_i\beta) \tag{1.59}$$

$$b_i = \frac{1}{2\sigma_{\varepsilon}^3}(1 - z_i)\,x_i\beta\,\widetilde{\lambda}\left(\frac{x_i\beta}{\sigma_{\varepsilon}}\right) - \frac{z_i}{2\sigma_{\varepsilon}^2} + \frac{z_i}{2\sigma_{\varepsilon}^4}(y_i - x_i\beta)^2 \tag{1.60}$$

avec $\widetilde{\lambda}(z) = \phi(z) / [1 - \Phi(z)] = \lambda(-z)$ et où la quantité z_i correspond à la variable dichotomique simple suivante :

$$z_i = \begin{cases} 1 & \text{si } y_i > 0 \\ 0 & \text{sinon} \end{cases} \tag{1.61}$$

Naturellement un estimateur de ce vecteur du gradient de la log-vraisemblance sous H_0 : $\alpha = 0$ peut être obtenu en remplaçant dans les expressions de a_i et de b_i les paramètres β et σ_ε^2 par leurs estimateurs du MV respectifs $\widehat{\beta}$ et $\widehat{\sigma}_\varepsilon^2$ obtenus sous l'hypothèse nulle $\alpha = 0$.

$$\left. \underset{(N,K+1+P)}{G\left(\widehat{\beta}, \widehat{\sigma}_\varepsilon^2, 0\right)} \right|_{\alpha=0} = \begin{pmatrix} \underset{(1,K+P+1)}{g_1\left(\widehat{\beta}, \widehat{\sigma}_\varepsilon^2, 0\right)} \\ \dots \\ \underset{(1,K+P+1)}{g_N\left(\widehat{\beta}, \widehat{\sigma}_\varepsilon^2, 0\right)} \end{pmatrix} = \begin{pmatrix} \widehat{a}_1 x_1' & \widehat{b}_1 & \widehat{\sigma}_\varepsilon^2 \widehat{b}_1 w_1' \\ & \dots & \\ \widehat{a}_N x_N' & \widehat{b}_N & \sigma_\varepsilon^2 \widehat{b}_N w_N' \end{pmatrix}$$

Reste alors à construire la matrice d'information de Fischer. Greene (1997) montre que sous H_0 l'inverse de la matrice d'information de Fischer peut s'écrire sous la forme :

$$\begin{aligned} \left. \underset{(K+P+1,K+P+1)}{I\left(\beta, \sigma_\varepsilon^2, \alpha\right)^{-1}} \right|_{\alpha=0} &= \underset{(K+P+1,N)}{G\left(\beta, \sigma_\varepsilon^2, 0\right)'} \underset{(N,K+P+1)}{G\left(\beta, \sigma_\varepsilon^2, 0\right)} \\ &= \sum_{i=1}^{N} \begin{pmatrix} a_i^2 x_i' x_i & a_i b_i x_i' & \sigma_\varepsilon^2 a_i b_i x_i' w_i \\ a_i b_i x_i & b_i^2 & \sigma_\varepsilon^2 b_i^2 w_i \\ \sigma_\varepsilon^2 a_i b_i w_i' x_i & \sigma_\varepsilon^2 b_i^2 w_i' & \sigma_\varepsilon^2 b_i^2 w_i' w_i \end{pmatrix} \end{aligned} \tag{1.62}$$

Un estimateur de la matrice d'information de Fischer est alors donné par :

$$\left. \widehat{I}\left(\widehat{\beta}, \widehat{\sigma}_\varepsilon^2, \alpha\right)^{-1} \right|_{\alpha=0} = \sum_{i=1}^{N} \begin{pmatrix} \widehat{a}_i^2 x_i' x_i & \widehat{a}_i \widehat{b}_i x_i' & \widehat{\sigma}_\varepsilon^2 \widehat{a}_i \widehat{b}_i x_i' w_i \\ \widehat{a}_i \widehat{b}_i x_i & \widehat{b}_i^2 & \widehat{\sigma}_\varepsilon^2 \widehat{b}_i^2 w_i \\ \widehat{\sigma}_\varepsilon^2 \widehat{a}_i \widehat{b}_i w_i' x_i & \widehat{\sigma}_\varepsilon^2 \widehat{b}_i^2 w_i' & \widehat{\sigma}_\varepsilon^2 \widehat{b}_i^2 w_i' w_i \end{pmatrix}$$

A partir de ces différents éléments on peut alors construire la statistique LM de la façon suivante :

Definition 1.16. ***Une autre expression de la statistique*** LM ***du test de l'hypothèse nulle d'homoscédasticité*** $H_0 : \alpha = 0$ ***est :***

$$\underset{(1,1)}{LM} = \underset{(1,N)}{e_N'} \underset{(N,K+P+1)}{G\left(\widehat{\beta}, \widehat{\sigma}_\varepsilon^2, 0\right)} \left[\underset{(K+P+1,N)}{G\left(\widehat{\beta}, \widehat{\sigma}_\varepsilon^2, 0\right)'} \underset{(N,K+P+1)}{G\left(\widehat{\beta}, \widehat{\sigma}_\varepsilon^2, 0\right)} \right]^{-1} \underset{(K+P+1,N)}{G\left(\widehat{\beta}, \widehat{\sigma}_\varepsilon^2, 0\right)'} \underset{(N,1)}{e_N} \tag{1.63}$$

où e_N ***désigne un vecteur unitaire de dimension*** $(N,1)$ ***et où*** $\widehat{\beta}$ ***et*** $\widehat{\sigma}_\varepsilon$ ***désignent les estimateurs du*** MV ***des paramètres*** β ***et*** σ_ε^2 ***obtenus sous l'hypothèse nulle*** $\alpha = 0$.

On peut montrer que cette expression de la statistique LM est identique à celle proposée dans la définition (1.15). La loi asymptotique et la règle de décision sont évidemment les mêmes que celles évoquées précédemment.

Il existe enfin une troisième façon d'obtenir la statistique LM :

Definition 1.17. ***Une autre expression de la statistique*** LM ***du test de l'hypothèse nulle d'homoscédasticité*** $H_0 : \alpha = 0$ ***est :***

$$LM = N\, R^2 \tag{1.64}$$

où N ***désigne le nombre d'oibservations et où*** R^2 ***est le coefficient de détermination de la régression du vecteur unitaire*** $e_N = (1, ...1)'$ ***de dimension*** $(N, 1)$ ***sur les*** $K+P+1$ ***colonnes de la matrice*** $G\left(\widehat{\beta}, \widehat{\sigma}_{\varepsilon}^2, 0\right)$.

En effet, une fois que l'on a construit la matrice $G\left(\widehat{\beta}, \widehat{\sigma}_{\varepsilon}^2, 0\right)$, on peut montrer que le coefficient de détermination de la régression de e_N sur les colonnes G_1, $G_2, .., G_{K+P+1}$ de cette matrice fournit au coefficient N près la valeur de la statistique LM. Si l'on pose $y = e_N$ et $X = G(.)$, on sait que le coefficient de la régression linéaire de y sur les colonnes de X est donné par la formule :

$$R^2 = \frac{1}{N}\left[y'X\left(X'X\right)^{-1}X'y\right]$$

où N désigne le nombre d'observation. On reconnaît ici la forme de la statistique LM au scalaire N près.

***** **

1°) Simulation : simuler biais à distance finie avec Hétéro à la Hurd.

2°) Application : Construire un test sous LIMDEP et application

***** **

1.6.2. Non normalité

La seconde principale hypothèse qui peut affecter de façon sensible les propriétés de l'estimateur du MV est l'hypothèse de non normalité des perturbations. Alors quelles sont le propriétés de l'estimateur du MV sous cette hypothèse de non normalité ? Nous admettrons le résultat suivant :

Proposition 1.18. ***De façon générale, on montre que l'estimateur du*** MV ***n'est pas convergent lorsque la vraie distribution des perturbations*** ε_i ***n'est pas normale.***

Goldberger (1980) a en effet démontré dans le cas d'un modèle Tobit simple tronqué, l'existence de biais asymptotique de l'estimateur du MV lorsque la vraie distribution des ε_i est une loi de Student, de Laplace ou une loi logistique. Pour démontrer ce résultat, Goldberger supposait que la variance des perturbations était toutefois connue. Arabmazar et Schmidt (1982) ont quant à eux montré que les biais étaient particulièrement accrus lorsque l'on levait cette hypothèse et que l'on supposait la variance des perturbations étaient inconnues. Intuitivement on conçoit qu'une erreur sur la distribution des perturbations et donc sur la forme de vraisemblance, peut conduire à l'apparition d'un biais dans les estimateurs du MV.

****

Effectuer Simulation Tobit Simple avec loi Student, de Laplace ou loi logistique

Estimation MV fonction de la censure et de N

****

Partant du résultat que l'application de la procédure du MV à des perturbations de loi non normale conduit à un biais asymptotique, il convient donc de proposer *un test permettant de repérer les cas où les perturbations du modèle sont distribuées selon une loi non normale.* Deux principales stratégie de test sont proposés pour détecter la non normalité des perturbations dans un modèle censuré ou tronqué :

1. *Une stratégie de test à la Hausman (1978) : Nelson (1981), Melenberg et Van Soest (1996)*

2. *Un test de spécification à la Hansen (1982) : Pagan et Vella (1989)*

Nous n'évoquerons ici que le premier type de test. Pour les tests de spécification reposant sur les conditions sur les moments voir Pagan et Vella (1989). **Considérons la démarche retenue par Melenberg et Van Soest (1996) fondée sur un test d'Hausman (1978) avec esstiamteur LAD de Powell (1984).**

Commençons par définir de façon général le principe d'un test de Hausman (1978). Ce test admet pour hypothèse nulle la normalité des résidus ε_i. Soit $\widehat{\beta}$ l'estimateur du MV du vecteur des paramètres β. On sait que cet estimateur est (i) convergent, (ii) asymptotiquement efficace et (iii) asymptotiquement biaisé sous l'hypothèse alternative H_1 de non normalité des perturbations.. Considérons un second estimateur, noté $\widetilde{\beta}$, du vecteur des paramètres β. On choisit cet estimateur de sorte à ce qu'il soit (i) moins efficace que l'estimateur du MV sous H_0 mais (ii) qu'il soit convergent sous H_0 et sous l'hypothèse alternative H_1. Il ne reste plus alors qu'à étudier la "distance" entre les deux estimateurs $\widehat{\beta}$ et $\widetilde{\beta}$. En effet :

- Si les deux estimateurs sont "proches" : cela signifie que les deux estimateurs sont non biaisés : l'hypothèse H_0 est acceptée

- Si les deux estimateurs sont suffisamment "éloignés" : cela signifie que l'estimateur $\widehat{\gamma}$ est biaisé : l'hypothèse H_0 est rejetée.

Reste alors à construire une mesure de la "distance" entre les deux estimateurs. Hausman dans son article de 1978 montre que de façon générale, sous ces hypothèses, la quantité définie converge une loi du Chi deux admettant pour degré de liberté le nombre de paramètre estimés sous l'hypothèse H_0 .

$$H_N = \left(\widehat{\beta} - \widetilde{\beta}\right)' V^{-1} \left(\widehat{\beta} - \widetilde{\beta}\right) \underset{N \to \infty}{\overset{L}{\longrightarrow}} \chi^2(K) \tag{1.65}$$

où $V = V\left(\widehat{\beta}\right) - V\left(\widetilde{\beta}\right)$ désigne la différence entre les matrices de variance covariances asymptotiques des deux estimateurs obtenues sous H_0. Ce qui est remarquable c'est que sous H_0, il n'est pas nécessaire de connaître les termes de covariances des deux estimateurs pour construire la statistique de test.

Proposition 1.19. ***Soit $\widehat{\beta}$ l'estimateur du MV du vecteur des paramètres β. Soit $\widehat{\beta}$ un estimateur convergent sous l'hypothèse de non des perturbations ε_i. Un test de l'hypothèse nulle de normalité peut être réalisé à partir de la statistique du test de Hausman (1978) :***

$$H_N = \left(\widehat{\beta} - \widetilde{\beta}\right)' \left[V\left(\widehat{\beta}\right) - V\left(\widetilde{\beta}\right)\right]^{-1} \left(\widehat{\beta} - \widetilde{\beta}\right) \tag{1.66}$$

où $V\left(\widehat{\beta}\right)$ et $V\left(\widetilde{\beta}\right)$ désignent les matrices de variance covariances asymptotiques des estimateurs sous H_0. Sous l'hypothèse nulle de normalité :

$$H_N \underset{N\to\infty}{\overset{L}{\longrightarrow}} \chi^2(K) \tag{1.67}$$

Ainsi, si la réalisation de la statistique H_N est supérieur au fractile à $\alpha\%$ de la loi du Chi-deux, la distance entre les deux estimateurs est grande : on rejette l'hypothèse nulle H_0 de normalité. L'estimateur du MV est alors asymptotiquement biaisé.

Pour construire ce test, reste à définir **un estimateur convergent du vecteur de paramètres β sous l'hypothèse de non normalité**. Il y a là aussi deux optiques :

1. *Soit on spécifie la distribution non normale des perturbations et l'on construit un estimateur du maximum de vraisemblance :* Amemiya et Boskin (1974) utilisent ainsi un estimateur du MV avec une distribution log-normale

2. *Soit on construit un estimateur convergent pour des formes très générales de distributions des perturbations à la fois normale et non normale :* Powell (1984), Melenberg et Van Soest (1996).

La première approche est à la fois risquée et relativement compliquée à mettre en oeuvre. Elle peut apparaître compliquée dans la mesure où pour certaines formes de distributions, il peut être délicat de construire la fonction de vraisemblance. De plus, rien ne garantit alors la concavité globale de cette fonction, ce qui peut poser des problèmes d'optimisation numérique. Mais elle de plus risquée dans la mesure où l'on rejette a priori la distribution normale pour spécifier une forme alternative de distribution : log-normale, Student, Laplace etc.. Or, rien ne garantit que les perturbations soient effectivement engendrées par cette distribution. Une erreur sur la forme de la distribution peut alors conduire à une évaluation biaisée des paramètres. C'est pourquoi, dans la littérature on privilégie généralement la seconde approche : *l'approche non paramétrique ou semi-paramétrique*. Melenberg et Van Soest (1996) proposent d'utiliser l'estiamteur LAD.

L'estimateur de Powell (1984) ou estimateur des Moindres Valeurs Absolues (LAD *Least Absolute Deviations*) est un exemple d'estimateur non paramétrique convergent sous l'hypothèse de non normalité.

Definition 1.20. ***L'estimateur des Moindres Valeurs Absolues (LAD Least Absolute Deviations) de Powell (1984) des paramètres β du modèle Tobit simple est défini par :***

$$\widehat{\beta}_P = \underset{\{\beta\}}{\arg\min} \left[\sum_{i=1}^{N} y_i - \max(0, x_i\beta)\right] \tag{1.68}$$

Powell montre que cet estimateur est asymptotiquement normal :

$$\sqrt{N}\left(\widehat{\beta}_P - \beta\right) \underset{N\to\infty}{\overset{L}{\longrightarrow}} N\left(0, V\left(\widehat{\beta}_P\right)\right) \tag{1.69}$$

où la matrice de variance covariance asymptotique est donnée par :

$$\underset{(K,K)}{V\left(\widehat{\beta}_P\right)} = 4f(0)^2 \lim_{N\to\infty} \frac{1}{N} \sum_{i:x_i\beta>0} x_i x_i' \tag{1.70}$$

où $f(.)$ désigne la fonction de densité des perturbations.

Paarsch (1984) a proposé différentes simulations de Monte Carlo de cet estimateur, de l'estimateur en deux étapes d'Heckman et de l'estimateur du MV obtenu sous l'hypothèse de normalité pour des modèles Tobit avec des distributions normal, exponentielle et de Cauchy. Sur de larges échantillons, l'estimateur de Powell est toujours meilleur (au sens du biais moyen) que l'estimateur d'Heckman et est meilleur que l'estimateur du MV dans le cas de distribution de Cauchy. *Powell (1984) note en outre que l'estimateur $\widehat{\beta}_P$ est convergent y compris sous l'hypothèse d'hétéroscédasticité.*

Ainsi, en utilisant la définition précédente du test d'Hausman et la définition de l'estimateur LAD on peut construire aisément un test particulier de l'hypothèse de non normalité fondé sur l'estimateur de Powell. Si l'on note $\widehat{\beta}$ l'estimateur du MV obtenu sous l'hypothèse de normalité et $\widehat{\beta}_P$ l'estimateur de Powell, la statistique du test d'Hausman devient :

$$H_N = \left(\widehat{\beta} - \widehat{\beta}_P\right)' \left[V\left(\widehat{\beta}\right) - V\left(\widehat{\beta}_P\right)\right]^{-1} \left(\widehat{\beta} - \widehat{\beta}_P\right) \tag{1.71}$$

où $V\left(\widehat{\beta}\right)$ et $V\left(\widehat{\beta}_P\right)$ désignent les matrices de variance covariances asymptotiques des estimateurs sous H_0 :

$$V\left(\widehat{\beta}_P\right) = 4f(0)^2 \lim_{N\to\infty} \frac{1}{N} \sum_{i:x_i\beta>0} x_i x_i' \tag{1.72}$$

$$V\left(\widehat{\beta}\right) = I(\beta)^{-1} \tag{1.73}$$

avec

$$I(\beta) = -E\left[\frac{\partial^2 \log L(\beta,\beta)}{\partial\beta\partial\beta'}\right] \tag{1.74}$$

Naturellement si le test conduit à rejeter l'hypothèse nulle de normalité, il convient de privilégier un estimateur convergent sous l'hypothèse de non normalité : l'estimateur LAD de Powell en est un, mais il existe de nombreux autres estimateur non paramétriques applicables dans ce cas.

******* ***************************

Application Eviews ou Limdep ou Matlab
******* ***************************

1.7. Extensions du modèle Tobit Simple : modèles à censure multiples

Différentes *extensions du modèle Tobit simple ont été proposées sans pour autant remettre en cause sa structure générale* : une variable dépendante correspondant à une variable latente observée sur un certain intervalle. Ces extensions portent finalement sur la définition de l'intervalle sur lequel est observé la variable latente y_i^*. En effet dans certaines applications, la variable dépendante peut être censurée à la fois à droite et à gauche. C'est par exemple le cas sur un marché de cotation où il existerait une limite inférieure et supérieure aux cours, auxquels cas la valeur du cours est fixée soit à un cours plancher soit à un cours plafond. On parle alors de **modèle Tobit à censures multiples**. Lorsque les seuils de censure sont identiques à tous les individus, on parle alors de **modèle Tobit à double censures**. Naturellement le modèle à double censure est un cas particulier du modèle à censures multiples, c'est pourquoi nous débuterons notre analyse par ce dernier.

1.7.1. Modèle Tobit simple à censures multiples

Le modèle Tobit simple à censures multiples s'écrit sous la forme suivante :

Definition 1.21. ***Un modèle Tobit simple à censures multiples est défini par :***

$$y_i = \begin{cases} c_{i,1} & si\ y_i^* \leq c_{i,1} \\ y_i^* & si\ c_{i,1} < y_i^* \leq c_{i,2} \\ c_{i,2} & si\ y_i^* \geq c_{i,2} \end{cases} \tag{1.75}$$

où $(c_{i,1}, c_{i,2}) \in \mathbb{R}^2$ ***désigne les bornes de censure et où :***

$$y_i^* = x_i\beta + \varepsilon_i \qquad \forall i = 1, ..N \tag{1.76}$$

où $x_i = \left(x_i^1..x_i^K\right), \forall i = 1, .., N$ ***désigne un vecteur de caractéristiques observables et où*** $\beta = (\beta_1...\beta_K)' \in \mathbb{R}^K$ ***est un vecteur de paramètres inconnus et où les perturbations*** ε_i ***sont distribués selon une loi*** $N\left(0, \sigma_\varepsilon^2\right)$***.***

Considérons un échantillon de N observations y_i, noté $y = (y_1, .., y_N)$. La la fonction de vraisemblance d'un modèle à censures multiples s'écrit sous la forme :

$$\begin{aligned} & L\left(y, \beta, \sigma_\varepsilon^2, c_{1,1}, c_{1,2}, ..., c_{N,1}, c_{N,2}\right) \\ = & \prod_{i:\, y_i = c_{i,1}} \Phi\left(\frac{c_{i,1} - x_i\beta}{\sigma_\varepsilon}\right) \prod_{i:\, y_i = c_{i,2}} \left[1 - \Phi\left(\frac{c_{i,2} - x_i\beta}{\sigma_\varepsilon}\right)\right] \prod_{i:\, y_i = y_i^*} \left(\frac{1}{\sigma_\varepsilon}\right) \phi\left(\frac{y_i - x_i\beta}{\sigma_\varepsilon}\right) \end{aligned} \tag{1.77}$$

Le premier terme désigne le produit des probabilités que les observations y_i prennent les valeurs de censures inférieures $c_{i,1}$:

$$Prob\,(y_i = c_{i,1}) = Prob\,(y_i^* \leq c_{i,1}) = \Phi\left(\frac{c_{i,1} - x_i\beta}{\sigma_\varepsilon}\right)$$

Le second terme désigne le produit des probabilités que les observations y_i prennent les valeurs de censures supérieures $c_{i,2}$:

$$Prob\,(y_i = c_{i,2}) = Prob\,(y_i^* \geq c_{i,2}) = \Phi\left(\frac{c_{i,2} - x_i\beta}{\sigma_\varepsilon}\right)$$

Enfin, le troisième terme représente tout simplement le produit des lois marginales des variables y_i lorsque ces dernières appartiennent à l'intervalle compris entre les deux bornes de censure. On sait que si $c_{i,1} < y_i^* \leq c_{i,2}$, on a part définition $y_i = y_i^* = x_i\beta + \varepsilon_i$ où les perturbations ε_i sont distribués selon une loi $N\left(0, \sigma_\varepsilon^2\right)$. On en déduit que les variables y_i sont alors distribuées selon une loi normale $N\left(x_i\beta, \sigma_\varepsilon^2\right)$. Ainsi, la loi marginale d'une observation y_i sur cet intervalle est définie par la quantité :

$$\frac{1}{\sigma_\varepsilon\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{y_i-x_i\beta}{\sigma_\varepsilon}\right)^2\right] = \left(\frac{1}{\sigma_\varepsilon}\right)\phi\left(\frac{y_i - x_i\beta}{\sigma_\varepsilon}\right)$$

où $\phi(.)$ désigne la fonction de densité associée à loi normale centrée réduite. On peut en déduire la log-vraisemblance dans un modèle à censures multiples :

Definition 1.22. ***La log-vraisemblance concentrée associée à un échantillon*** $y = (y_1, .., y_N)$ ***dans un modèle Tobit simple à censures multiples s'écrit :***

$$\begin{aligned}
&\log L\left(y, \beta, \sigma_\varepsilon^2, c_{1,1}, c_{1,2}, ..., c_{N,1}, c_{N,2}\right) \\
= \quad &\sum_{i:\, y_i = c_{i,1}} \log\left[\Phi\left(\frac{c_{i,1} - x_i\beta}{\sigma_\varepsilon}\right)\right] + \sum_{i:\, y_i = c_{i,2}} \log\left[1 - \Phi\left(\frac{c_{i,2} - x_i\beta}{\sigma_\varepsilon}\right)\right] \\
&-\frac{N_1}{2}\log\left(\sigma_\varepsilon^2\right) - \frac{1}{2\sigma_\varepsilon^2}\sum_{i:\, y_i = y_i^*}\left(y_i - x_i\beta\right)^2 \qquad (1.78)
\end{aligned}$$

où N_1 ***désigne le nombre d'observations pour lesquelles*** $y_i = y_i^*$.

Il est souvent utile dans ces modèles de déterminer les espérances conditionnelles de la variable dépendante limitée et de la variable non censurée, notamment pour calculer les effets marginaux. On pose $\forall i = 1, ..N$:

$$\Phi_{1,i} = \Phi\left(\frac{c_{i,1} - x_i\beta}{\sigma_\varepsilon}\right) \qquad \Phi_{2,i} = \Phi\left(\frac{c_{2,1} - x_i\beta}{\sigma_\varepsilon}\right)$$

Dès lors, on montre (Alban 2000) que l'espérance de la variable dépendante limitée est :

$$E\left(y_i / x_i, c_{i,1} < y_i^* \leq c_{i,2}\right) = x_i\beta + \sigma_\varepsilon\left(\frac{\phi_{1,i} - \phi_{2,i}}{\Phi_{2,i} - \Phi_{1,i}}\right) \qquad (1.79)$$

En effet dans le cas du modèle Tobit simple on a uniquement une censure à gauche, ce qui se traduit par des seuils de censure égaux à $c_{1,i} = 0$ et $c_{2,i} = +\infty$. On obtient alors :

$$\phi_{1,i} = \phi\left(\frac{-x_i\beta}{\sigma_\varepsilon}\right) \qquad \phi_{2,i} = \lim_{c_{2,i}\to\infty}\phi\left(\frac{c_{2,1} - x_i\beta}{\sigma_\varepsilon}\right) = 0$$

$$\Phi_{1,i} = \Phi\left(\frac{-x_i\beta}{\sigma_\varepsilon}\right) \qquad \Phi_{2,i} = \lim_{c_{2,i}\to\infty}\Phi\left(\frac{c_{2,1} - x_i\beta}{\sigma_\varepsilon}\right) = 1$$

Ainsi, on montre que l'espérance conditionnelle se ramène à l'expression suivante dans le cas du modèle Tobit simple :

$$\begin{aligned}
E\left(y_i / x_i, c_{i,1} < y_i^* \leq c_{i,2}\right) &= x_i\beta + \sigma_\varepsilon\left(\frac{\phi_{1,i}}{1 - \Phi_{1,i}}\right) \\
&= x_i\beta + \sigma_\varepsilon\frac{\phi\left(x_i\beta / \sigma_\varepsilon\right)}{\Phi\left(x_i\beta / \sigma_\varepsilon\right)} \\
&= x_i\beta + \sigma_\varepsilon\lambda\left(\frac{-x_i\beta}{\sigma_\varepsilon}\right)
\end{aligned}$$

On retouve ainsi l'expression standard d'une espérance conditionnelle d'une loi normale censurée.

Revenons au cas général, c'est à dire au cas du modèle Tobit à censurs multiples. L'espérance de la variable dépendante non censurée est alors :

$$\begin{aligned} E\left(y_i/x_i\right) &= Prob\left(y_i = c_{i,1}\right) \times c_{i,1} + Prob\left(y_i = c_{i,2}\right) \times c_{i,2} \\ &\quad + Prob\left(c_{i,1} < y_i^* \leq c_{i,2}\right) \times E\left(y_i/x_i, c_{i,1} < y_i^* \leq c_{i,2}\right) \end{aligned}$$

On obtient alors la formule suivante :

$$\begin{aligned} E\left(y_i/x_i\right) &= \Phi_{1,i}\, c_{i,1} + \left(1 - \Phi_{2,i}\right)\, c_{i,2} + x_i\beta\left(\Phi_{2,i} - \Phi_{1,i}\right) \\ &\quad + \sigma_\varepsilon\left(\phi_{1,i} - \phi_{2,i}\right) \end{aligned} \tag{1.80}$$

1.7.2. Modèle Tobit simple à double censure : Rosett et Nelson (1975)

Comme nous l'avons dit précédemment le modèle Tobit simple à double censure est un cas particulier du modèle Tobit simple à censures multiples. C'est un modèle dans lequel on suppose que les seuils de censure à droite et gauche sont identiques pour tous les individus.

$$c_{i,1} = c_1 \qquad c_{i,2} = c_2 \quad \forall i = 1, ..N \tag{1.81}$$

Definition 1.23. ***Un modèle Tobit simple à double censure (modèle de friction ou de Rosett) est un modèle où les seuils de censures à gauche et à droite sont identiques pour tous les individus.***

$$y_i = \begin{cases} c_1 & si\ y_i^* \leq c_1 \\ y_i^* & si\ c_1 < y_i^* \leq c_2 \\ c_2 & si\ y_i^* \geq c_2 \end{cases} \tag{1.82}$$

où $(c_1, c_2) \in \mathbb{R}^2$.

Ce modèle est aussi parfois appelé **modèle de Rosett** ou **modèle de friction** du fait de l'application proposée par cet auteur. C'est en effet Rosett et Nelson (1975) qui ont proposé la première modélisation Tobit simple à double censure. La log-vraisemblance concentrée associée à un échantillon $y = (y_1, .., y_N)$ dans un modèle Tobit simple à double censure s'écrit :

$$\begin{aligned} &\log L\left(y, \beta, \sigma_\varepsilon^2, c_1, c_2\right) \\ =& \sum_{i:\, y_i = c_1} \log\left[\Phi\left(\frac{c_1 - x_i\beta}{\sigma_\varepsilon}\right)\right] + \sum_{i:\, y_i = c_2} \log\left[1 - \Phi\left(\frac{c_2 - x_i\beta}{\sigma_\varepsilon}\right)\right] \\ & - \frac{N_1}{2}\log\left(\sigma_\varepsilon^2\right) - \frac{1}{2\sigma_\varepsilon^2}\sum_{i:\, y_i = y_i^*}\left(y_i - x_i\beta\right)^2 \end{aligned} \tag{1.83}$$

où N_1 désigne le nombre d'observations pour lesquelles $y_i = y_i^*$.

Ce modèle est utilisé dans des applications où la variable dépendante ne répond qu'à de fortes variations (ou de fortes valeurs) des variables explicatives. Nous allons à présent évoquer deux exemples de modèle Tobit simples à double censure :

1. Modèle de distributions de dividendes : Maddala (1977)
2. Modèle d'investissement financier avec coût de transaction : Rosett (1959)

Commençons par le modèle de Maddala (1977).

Politique de Dividendes : Maddala (1977) Maddala (1977) remarque que le modèle Tobit simple à double censure ou modèle de friction est particulièrement adapté pour modéliser la politique de dividendes des entreprises, les variations des salaires offerts par les firmes ou *tout autre décisions pour lesquelles les firmes répondent par saut après un certain effort cumulatif.*

Considérons une société par action qui distribue des dividendes à ses actionnaires. Soit y_t^* le montant désiré de dividende qui dépend d'un ensemble de caractéristiques de l'entreprise : montant des bénéfices, décisions d'autofinancement, montant des investissements etc..On pose que ces caractéristiques peuvent être représentées par un vecteur x_t et qu'elles sont liées au dividende potentiel par une relation du type $y_t^* = x_t\beta + \varepsilon_t$ où ε_t est distribué selon une loi $N\left(0, \sigma_\varepsilon^2\right)$. Si l'on suppose que le mécanisme de distribution des dividendes est coûteux, on suppose que l'entreprise limite leur distribution dans le temps (au plus une fois par an) mais aussi suivant leur montant :

- L'entreprise ne verse des dividendes que si le montant potentiel de ces dividendes est supérieur à un certain seuil c_1
- L'entreprise limite le montant de ces dividendes à un niveau c_2 afin de maintenir une marge de manoeuvre financière.

Dès lors, si l'on note y_t le montant des dividendes effectivement versés on a :

$$y_t = \begin{cases} 0 & \text{si } y_t^* \leq c_1 \\ y_t^* & \text{si } c_1 < y_t^* \leq c_2 \\ c_2 & \text{si } y_t^* \geq c_2 \end{cases}$$

Investissements financiers et coûts de transaction : Rosett (1959) Rosett (1959) avait déjà proposé une application dans laquelle apparaissait une version particulière du modèle Tobit simple à double censure. Ce n'est qu'en 1975 que Rosett et Nelson donneront la forme générale du modèle, mais c'est pourquoi le nom de modèle de Rosett ou modèle de friction est généralement attribué à ce modèle.

Dans son application Rosett(1959) considère un modèle d'investissement dans des actifs financiers où les coûts de transaction peuvent limiter le volume des transactions par rapport au niveau désiré. Le modèle suppose que les modifications dans la position de l'investisseur, c'est à dire la décision d'achat ou de vente, dépend des variations du rendement. Soit y_t^* la variation désirée de la position du titre (montant acheté ou vendu), y_t la variation de la position effective et x_t la variation du rendement du titre. L'investisseur n'effectuera une transaction que si les variations du rendement sont suffisamment importantes. La position réelle du titre ne change donc pas pour de petites variations à la hausse ou à la baisse du rendement et donc de la position désirée. Supposons que la position désirée soit liée à la variation du rendement par la relation $y_t^* = x_t\beta + \varepsilon_t$ où ε_t est distribué selon une loi $N\left(0, \sigma_\varepsilon^2\right)$. Le modèle est alors défini par :

$$y_t = \begin{cases} y_t^* - c_1 & \text{si } y_t^* \leq c_1 \\ 0 & \text{si } c_1 < y_t^* \leq c_2 \\ y_t^* - c_2 & \text{si } y_t^* \geq c_2 \end{cases}$$

où c_1 est le niveau de baisse de la position désiré déclenchant la vente et $c_2 > 0$ le niveau déclenchant l'achat.

1.7.3. Application modèle à double censure

****

Application Eviews ou Limdep
****

2. Les Modèles Tobit Généralisés

Nous allons à présent envisager des modèles Tobit incluant au moins deux variables y_1 et y_2. Cette classe de modèle s'appelle **la classe des modèles Tobit généralisés**. Comme nous l'avons dit en introduction, Amemiya (1983) propose de classer les différents modèle Tobit généralisés en 5 principales classes. Le modèle Tobit simple étant par convention défini comme le modèle Tobit de type I. L'auteur propose la classification suivante en fonction de la forme de la vraisemeblance et des propriétés des variables introduites dans le modèle :

Tableau 2.1: Fonctions de Vraisemblance des Modèles Tobit Généralisés

Modèle	Forme de la Vraisemblance	y_1	y_2	y_3
Tobit Type I	$P(y_1 < 0) \times P(y_1)$	C	—	—
Tobit Type II	$P(y_1 < 0) \times P(y_1 > 0, y_2)$	D	C	—
Tobit Type III	$P(y_1 < 0) \times P(y_1, y_2)$	C	C	
Tobit Type IV	$P(y_1 < 0, y_3) \times P(y_1, y_2)$	C	C	C
Tobit Type V	$P(y_1 < 0, y_3) \times P(y_1 > 0, y_2)$	D	C	C

Source : Amemiya (1983), Tables 1 et 2, page 30, D : variable dichotomique, C : censurée

Ainsi dans le cas du modèle standard, la notation $P(y_1 < 0) \times P(y_1)$ d'Amemiya désigne une fonction de vraisemblance de la forme $\prod_{y_i=0} P\left(y_{1,i}^* \leq 0\right) . \prod_{y_i>0} f(y_{1,i})$ où $f(y_{1,i})$ désigne la densité marginale de la variable $y_{1,i}$ distribuée selon une loi $N\left(x_i\beta, \sigma_\varepsilon^2\right)$. Les notations pour les autres modèles sont similaires, sachant que $P(y_1, y_2)$ désigne la densité jointe des variables y_1 et y_2. L'autre façon de distinguer les différents modèles Tobit consiste à distinguer les propriétés des variables du système en différentiant les variables dichotomiques D et les variables censurées C. En effet dans ces modèles on a deux types de modélisation de la variable y_1 :

1. Soit le signe de la variable y_1 (par exemple dans le Tobit II) conditionne la modélisation (la censure ou la troncation) d'une autre variable : on a alors une modélisation dichotomique D sur cette variable.

2. Soit la variable y_1 joue un double rôle : son signe détermine le modèle (la censure ou la troncation) d'une autre variable mais elle est en outre elle même une variable censurée.

Nous allons dans un premier temps nous intéresser au **modèle Tobit généralisé de type II** qui est très souvent utilisé et dont la structure est très similaire à celle des **modèles de selection** (ou *modèles à troncature auxiliaire*) popularisés par Heckman et Gronau. Enfin, nous étudierons plus succintement les autres modèles Tobit généralisés recencés par Amemiya (1983).

2.1. Modèle Tobit Généralisé Type 2

Reprenons l'exemple des dépenses de consommation en biens durables. Dans le modèle Tobit simple (ou modèle Tobit de type I), nous avons supposé que le consommateur décide simultanément (i) du fait qu'il va ou non consommer et (ii) du montant de revenu qu'il va affecter à cette consommation. Un modèle alternatif consisteriat à supposer un comportement séquentiel. Dans une première étape l'individu décide ou non de consommer : cette décision peut être représentée par un modèle qualititatif dichotomique basée sur un certain critère $y_{1,i}^*$.

$$\begin{cases} \text{si } y_{1,i}^* > 0 \text{ l'individu } i \text{ décide de consommer} \\ \text{si } y_{1,i}^* \leq 0 \text{ l'individu } i \text{ décide de ne pas consommer} \end{cases}$$

Dans une seocnde étape, s'il a décidé de consommer, l'individu décide du montant qu'il va consacrer à l'achat du bien. On a alors un modèle de données censurées puisque, si l'on note $y_{2,i}$ la consommation effective de l'agent i, celle-ci est définie par $\forall i = 1, ..N$:

$$y_{2,i} = \begin{cases} y_{2,i}^* & \text{si } y_{1,i}^* > 0 \\ 0 & \text{si } y_{1,i}^* \leq 0 \end{cases} \tag{2.1}$$

Cette formulation généralise le modèle Tobit simple dans la mesure om l'on retouve le modèle Tobit simple en posant $y_{1,i,}^* = y_{2,i}^*$. L'avantage de cette modélisation est qu'elle permet notamment de faire apparaître la plus ou moins forte corrélation pouvant exister entre les deux décisions (i) décision de consommation (ii) décision du montant consommé. On a bien un **modèle Tobit généralisé de type II** puisque seul le signe de la variable $y_{1,i}^*$ représenté par la variable dichotomique $y_{1,i} = \mathbb{I}\left(y_{1,i}^* > 0\right)$ importe ($y_{1,i}$ est une variable D) tandis que la variable y_2 est censurée ($y_{2,i}$ est une variable C).

2.1.1. Définition du Tobit généralisé de type II

Ainsi, un modèle Tobit généralisé de type II est définie de la façon suivante :

Definition 2.1. ***Un modèle Tobit généralisé de type II est défini par*** $\forall i = 1, ..N$ ***:***

$$y_{2,i} = \begin{cases} y_{2,i}^* & si\ y_{1,i}^* > 0 \\ 0 & si\ y_{1,i}^* \leq 0 \end{cases} \tag{2.2}$$

$$y_{1,i}^* = x_{1,i}\beta_1 + \varepsilon_{1,i} \tag{2.3}$$

$$y_{2,i}^* = x_{2,i}\beta_2 + \varepsilon_{2,i} \tag{2.4}$$

où $x_{j,i} = \left(x_{j,i}^1..x_{j,i}^{K_j}\right)$ ***avec*** $j = 1, 2$ ***désignent deux vecteurs de caractéristiques observables, où les vecteurs*** $\beta_j = \left(\beta_{j,1}...\beta_{j,K_j}\right)' \in \mathbb{R}^{K_j}$, $j = 1, 2$ ***sont des vecteurs de paramètres inconnus et où les perturbations*** $\varepsilon_{j,i}$ ***sont distribués selon une loi*** $N\left(0, \sigma_j^2\right)$ ***,*** $j = 1, 2$ ***avec*** $E\left(\varepsilon_{1,i}\varepsilon_{2,i}\right) = \sigma_{12}$, $\forall i = 1, ..N$.

Ainsi, seul le signe de la variable $y_{1,i}^*$ est observable et la variable $y_{2,i}^*$ est observable uniquement lorsque $y_{1,i}^* > 0$. On suppose que les variables $x_{1,i}$ sont observables pour tous les individus de l'échantillon, tandis qu'il n'est pas nécessaire que les variables $x_{2,i}$ soient observables pour les individus pour lesquels $y_{1,i}^* \leq 0$. Par la suite, on supposera tout de même que ces caractéristiques sont observables pour tous les individus, ce qui confère un statut de variable censurée à la variable $y_{2,i}$. Enfin, pour simplifier les notations, on introduit la variable dichotomique $z_{1,i}$ telle que :

$$z_{1,i} = \begin{cases} 1 & \text{si } y_{1,i}^* > 0 \\ 0 & \text{si } y_{1,i}^* \leq 0 \end{cases} \tag{2.5}$$

En d'autres termes, *les couples de variables* $(z_{1,i}, y_{2,i})$ *constituent les variables dépendantes observées du système.* Il convient en outre de noter la propriété suivante :

Remark 4. *Contrairement au cas du modèle Tobit simple, dans un modèle Tobit généralisé de type II, la variable dépendante* $y_{2,i}$ *peut prendre des valeurs négatives.*

Une telle propriété peut dans certains problèmes économiques être génante, c'est pourquoi Cragg (1971) a proposé des modèles qui assure la non-négativité de $y_{2,i}$.

2.1.2. Estimation par Maximum de Vraisemblance

Naturellement, tout comme dans le cas du modèle Tobit simple, les paramètres du modèle Tobit généralisé peuvent être estimés par maximum de Vraisemblance (MV). Commençons par définir la vraisemblance dans un tel modèle. D'après la forme générique donnée par Amemiya (1983), pour le modèle de type II, on a une vraisemblance de la forme $P(y_1 < 0) \times P(y_1 > 0, y_2)$. On pose $\theta = \left(\beta_1, \beta_2, \sigma_1^2, \sigma_2^2, \sigma_{12}\right)$ l'ensemble des paramètres du modèle. Formellement, si l'on considère un échantillon $y_2 = (y_{2,1}, .., y_{2,N})$ et un ensemble d'observations $z_1 = (z_{1,1}, .., z_{1,N})$, la vraisemblance s'écrit sous la forme :

$$L(y_2, z, \theta) = \prod_{i:\, y_{2,i}=0} Prob\left(y_{1,i}^* \leq 0\right) \prod_{i:\, y_{2,i}=y_{2,i}^*} f\left(y_{2,i} / \, y_{1,i}^* > 0\right) Prob\left(y_{1,i}^* > 0\right) \tag{2.6}$$

où $f\left(y_{2,i} / \, y_{1,i}^* > 0\right)$ désigne la densité conditionnelle de $y_{2,i}$ sachant $y_{1,i}^* > 0$. Réécrivons la seconde partie de cette fonction. On note $f_{y_1}(.)$ la fonction de densité marginale associée à $y_{1,i}^*$, il vient :

$$Prob\left(y_{1,i}^* > 0\right) = \int_0^\infty f_{y_1}(z)\, dz$$

Dès lors, on peut érécrire le second membre de la fonction de vraisemblance sous la forme d'une intégrale simple définie sur la fonction de densité jointe des variables $y_{1,i}^*$ et $y_{2,i}^*$, notée $f_{y_1,y_2}(.,.)$. En effet, en omettant les indices i pour simplifier les notations, il vient :

$$\begin{aligned} f\left(y_2 / \, y_1^* > 0\right) Prob\left(y_1^* > 0\right) &= \int_0^\infty f\left(y_2 / \, z\right) \, f_{y_1}(z)\, dz \\ &= \int_0^\infty f_{y_1,y_2}(y_2, z) \, dz \qquad (2.7) \\ &= \int_0^\infty f_{y_1,y_2}\left(y_2, y_1^*\right) \, dy_1^* \qquad (2.8) \end{aligned}$$

Car en effet, on a par définition $f\left(y_2 / \, z\right) \, f_{y_1}(z) = f_{y_1,y_2}(y_2, z)$. *Toute l'astuce consiste alors à réécrire la densité jointe* $f_{y_1,y_2}\left(y_2, y_1^*\right)$ *en fonction de la densité conditionnelle de* y_1^* *par rapport à* y_2. En effet, on peut écrire cette quantité sous la forme suivante :

$$f_{y_1,y_2}\left(y_2, y_1^*\right) = f\left(y_1^* / \, y_2\right) \, f_{y_2}(y_2)$$

où $f_{y_2}(.)$ la densité marginale de la variable y_2^*. On obtient ainsi l'expression suivante :

$$\begin{aligned} f\left(y_2/\, y_1^*>0\right) Prob\left(y_1^*>0\right) &= \int_0^\infty f\left(y_1^*/\, y_2\right) f_{y_2}\left(y_2\right) dy_1^* \\ &= f_{y_2}\left(y_2\right) \int_0^\infty f\left(y_1^*/\, y_2\right) dy_1^* \end{aligned} \tag{2.9}$$

Quel est l'avantage de cette expression ? Cette expression fait apparaître la densité conditionnelle de y_1^* sachant que $y_2^* = y_2$, notée $f\left(y_1^*/\, y_2\right)$, qu'il est relativement facile de calculer. On utilise pour cela le résultat suivant :

Proposition 2.2. ***Soit*** (y_1, y_2) ***un couple de v.a.r. distribuées selon des lois normales respectives*** $N\left(\mu_1, \sigma_1^2\right)$ ***et*** $N\left(\mu_2, \sigma_2^2\right)$, ***telles que*** $E\left(y_1 y_2\right) = \sigma_{12}$, ***la loi conditionnelle de*** y_1 ***sachant que*** $y_2 = \overline{y}_2$, ***est une loi normale d'espérance*** $E\left(y_1/\, y_2 = \overline{y}_2\right)$ ***avec***

$$E\left(y_1/\, y_2 = \overline{y}_2\right) = \mu_1 + \frac{\sigma_{12}}{\sigma_2^2}\left(\overline{y}_2 - \mu_2\right) \tag{2.10}$$

et de variance $V\left(y_1/\, y_2 = \overline{y}_2\right)$ ***avec :***

$$V\left(y_1/\, y_2 = \overline{y}_2\right) = \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2} \tag{2.11}$$

Ainsi, dans le cadre du modèle Tobit, la loi conditionnelle de y_1^* sachant que $y_2^* = y_2$, est une loi normale d'espérance $E\left(y_1^*/\, y_2 = y_2\right)$ et de variance $V\left(y_1^*/\, y_2 = y_2\right)$ avec

$$E\left(y_1^*/\, y_2^* = y_2\right) = x_1\beta_1 + \frac{\sigma_{12}}{\sigma_2^2}\left(y_2 - x_2\beta_2\right) \tag{2.12}$$

$$V\left(y_1^*/\, y_2^* = y_2\right) = \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2} \tag{2.13}$$

On en déduit la relation suivante :

$$\begin{aligned} \int_0^\infty f\left(y_1^*/\, y_2\right) dy_1^* &= 1 - \int_{-\infty}^0 f\left(y_1^*/\, y_2\right) dy_1^* \\ &= 1 - \Phi\left(\frac{0 - E\left(y_1^*/\, y_2^* = y_2\right)}{\sqrt{V\left(y_1^*/\, y_2^* = y_2\right)}}\right) \\ &= \Phi\left(\frac{E\left(y_1^*/\, y_2^* = y_2\right)}{\sqrt{V\left(y_1^*/\, y_2^* = y_2\right)}}\right) \end{aligned}$$

On en déduit finallement que :

$$\int_0^\infty f\left(y_1^*/\, y_2\right) dy_1^* = \Phi\left[\frac{x_1\beta_1 + \left(\sigma_{12}/\sigma_2^2\right)\left(y_2 - x_2\beta_2\right)}{\sqrt{\sigma_1^2 - \left(\sigma_{12}^2/\sigma_2^2\right)}}\right] \tag{2.14}$$

où $\Phi(.)$ désigne la fonction de répartition de la loi $N(0,1)$. Sachant que y_2^* suit une loi $N\left(x_2\beta_2, \sigma_2^2\right)$, on montre alors que :

$$f_{y_2}\left(y_2\right) \int_0^\infty f\left(y_1^*/\, y_2\right) dy_1^* = \frac{1}{\sigma_2}\phi\left(\frac{y_2 - x_2\beta_2}{\sigma_2}\right) \Phi\left[\frac{x_1\beta_1 + \left(\sigma_{12}/\sigma_2^2\right)\left(y_2 - x_2\beta_2\right)}{\sqrt{\sigma_1^2 - \left(\sigma_{12}^2/\sigma_2^2\right)}}\right]$$

Sachant que $f\left(y_2/\,y_1^*>0\right)Prob\left(y_1^*>0\right)=\int_0^\infty f\left(y_1^*/\,y_2^*\right)\,f_{y_2}\left(y_2^*\right)dy_2^*$, on montre donc finallement que :

$$f\left(y_2/\,y_1^*>0\right)Prob\left(y_1^*>0\right)=\frac{1}{\sigma_2}\phi\left(\frac{y_2-x_2\beta_2}{\sigma_2}\right)\,\Phi\left[\frac{x_1\beta_1+\left(\sigma_{12}/\sigma_2^2\right)\left(y_2-x_2\beta_2\right)}{\sqrt{\sigma_1^2-\left(\sigma_{12}^2/\sigma_2^2\right)}}\right]$$

Il est alors immédiat d'écrire la vraisemblance du modèle Tobit généralisé à partir de l'équation (2.6).

Definition 2.3. ***La vraisemblance associée à un échantillon*** $y_2=\left(y_{2,1},..,y_{2,N}\right)$ ***et un ensemble d'observations*** $z_1=\left(z_{1,1},..,z_{1,N}\right)$ ***dans un modèle Tobit généralisé de type II de paramètres*** $\theta=\left(\beta_1,\beta_2,\sigma_1^2,\sigma_2^2,\sigma_{12}\right)$ ***s'écrit :***

$$\begin{aligned}L\left(y_2,z,\theta\right) &= \prod_{i:y_{2,i}=0}\left[1-\Phi\left(\frac{x_{1,i}\beta_1}{\sigma_1}\right)\right] \\ &\quad\times\prod_{i:y_{2,i}=y_{2,i}^*}\frac{1}{\sigma_2}\phi\left(\frac{y_2-x_2\beta_2}{\sigma_2}\right)\,\Phi\left[\frac{x_1\beta_1+\left(\sigma_{12}/\sigma_2^2\right)\left(y_2-x_2\beta_2\right)}{\sqrt{\sigma_1^2-\left(\sigma_{12}^2/\sigma_2^2\right)}}\right]\end{aligned}\tag{2.15}$$

On constate que cette fonction de vraisemblance ne dépend de σ_1 que par l'intermédiaire du ratio β_1/σ_1 : il y a donc un **problème d'identifiabilité**. Seules sont identifiables les fonctions de σ_{12}, σ_2, β_2 et β_1/σ_1. On peut donc sans perte de généralité poser $\sigma_1=1$ ce qui permet d'identifier le reste des paramètres. Dans le cas où β_1 et β_2 comportent des éléments en commun, le paramètre σ_1 est de nouveau identifiable.

De la même façon que pour le modèle tobit simple, la fonction de vraisemblance associée à un modèle tobit généralisée n'est pas globalement concave. C'est pourquoi, il est souvent intéressant d'utiliser la re-parmétrisation d'Olsen :

$$h_2=\frac{1}{\sigma_2}\qquad\theta_1=\frac{\beta_1}{\sigma_1}\qquad\theta_2=\frac{\beta_2}{\sigma_2}\tag{2.16}$$

Si l'on note ρ la corrélation entre les deux perturbations telle que :

$$\sigma_{12}=\rho\,\sigma_1\,\sigma_2\tag{2.17}$$

alors on montre le résultat suivant.

Proposition 2.4. ***Pour un niveau de corrélation*** ρ ***des chocs donnés, la log-vraisemblance d'un modèle tobit généralisé re-paramétré en*** $h_2=\sigma_2^{-1}$, $\theta_1=\beta_1/\sigma_1$ ***et*** $\theta_2=\beta_2/\sigma_2$ ***est globalement concave :***

$$\begin{aligned}\log L\left(y_2,z,h_2,\theta_1,\theta_2,\rho\right) &= \sum_{i:y_{2,i}=0}\log\left[1-\Phi\left(x_{1,i}\theta_1\right)\right]+N_1\log\left(h_2\right) \\ &\quad+\sum_{i:y_{2,i}=y_{2,i}^*}\log\left(h_2y_2-x_2\theta_2\right) \\ &\quad+\sum_{i:y_{2,i}=y_{2,i}^*}\log\Phi\left[\frac{x_1\theta_1+\rho\left(h_2y_2-x_2\theta_2\right)}{\sqrt{1-\rho^2}}\right]\end{aligned}$$

Ainsi il peut être intéressant de modifier la méthode du maximum de vraisemblance pour tenir compte de la concavité partielle de la log-vraisemblance en h_2, θ_1 et θ_2 à ρ fixé. On peut par exemple faire un balayage sur ρ et maximiser pour chauqe valeur de ρ retenue la vraisemblance par rapport à h_2, θ_1 et θ_2. Il n'y a alors aucun problème de choix dans les conditions initiales en raison de la propriété de concavité. Puis dans un second temps, on retient la valeur de ρ qui maximise la valeur de la vraisemblance et l'on en déduit les estimateurs correspondant des autres paramètres.

2.1.3. Estimation en deux étapes : Heckman (1976)

Généralement les paramètres des modèles Tobit généralisés sont estimés par MV. Toutefois, il peut être utile de recourir à d'autres méthodes d'estimation simples, qui même si elles ne sont pas efficaces, permettent d'avoir une première idée de l'échelle de grandeur des paramètres et qui peuvent en outre servir dans les phases de détermination des conditions initiales dans les algorithmes d'optimisation numérique de la vraisemblance. Parmi ces méthodes d'estimation simples, on retrouve bien évidemment la méthode d'estimation en deux étapes proprosée par Heckman (1976) et présentée précédemment dans le cas du modèle Tobit simple.

On considère le modèle suivant :

$$y_{2,i} = \begin{cases} y_{2,i}^* & \text{si } y_{1,i}^* > 0 \\ 0 & \text{si } y_{1,i}^* \leq 0 \end{cases} \tag{2.18}$$

$$y_{1,i}^* = x_{1,i}\beta_1 + \varepsilon_{1,i} \tag{2.19}$$

$$y_{2,i}^* = x_{2,i}\beta_2 + \varepsilon_{2,i} \tag{2.20}$$

où les perturbations $\varepsilon_{j,i}$ sont distribués selon une loi $N\left(0, \sigma_j^2\right)$, $j = 1, 2$ avec $E\left(\varepsilon_{1,i}\varepsilon_{2,i}\right) = \sigma_{12}$, $\forall i = 1, ..N$. Pour construire l'estimateur en deux étapes d'Heckman, on cherche tout d'abord à construire l'espérance conditionnelle $E\left(y_{2,i}^* / \, y_{1,i}^* > 0\right)$. Pour cela, considérons l'-expression de $y_{2,i}^*$ et exprimons là en fonction de la projectioon linéire des résidus $\varepsilon_{2,i}$ sur les résidus $\varepsilon_{1,i}$. Compte tenu des hypothèses faites sur les pertubations, on a par hypothèse $\varepsilon_{2,i} = \left(\sigma_{12}\sigma_1^{-2}\right)\varepsilon_{1,i} + \mu_{2,i}$ où $\mu_{2,i}$ est indépendant de $\varepsilon_{1,i}$ et normallement distribué de moyenne nulle de de variance égale à $\sigma_2^2 - \sigma_{12}^2\sigma_1^{-1}$. Ainsi, on obtient :

$$y_{2,i}^* = x_{2,i}\beta_2 + \varepsilon_{2,i} = x_{2,i}\beta_2 + \sigma_{12}\sigma_1^{-2}\left(y_{1,i}^* - x_{1,i}\beta_1\right) + \mu_{2,i}$$

On en déduit alors l'expression $E\left(y_{2,i}^* / \, y_{1,i}^* > 0\right)$ en fonction de celle de $E\left(y_{1,i}^* / \, y_{1,i}^* > 0\right)$ que l'on avait déjà construit dans le chapitre précédent. En effet, sachant :

$$E\left(y_{1,i}^* / \, y_{1,i}^* > 0\right) = x_{1,i}\beta_1 + \sigma_1\lambda\left(x_{1,i}\theta_1\right)$$

on montre que :

$$\begin{aligned} E\left(y_{2,i}^* / \, y_{1,i}^* > 0\right) &= x_{2,i}\beta_2 + \sigma_{12}\sigma_1^{-2}\left[E\left(y_{2,i}^* / \, y_{1,i}^* > 0\right) - x_{1,i}\beta_1\right] + \mu_{2,i} \\ &= x_{2,i}\beta_2 + \sigma_{12}\sigma_1^{-1}\lambda\left(x_{1,i}\theta_1\right) + \mu_{2,i} \end{aligned}$$

avec $\mu_{2,i} = \varepsilon_{2,i} - \left(\sigma_{12}\sigma_1^{-2}\right)\varepsilon_{1,i}$.

Ainsi, pour les observations $y_{2,i}$ positives, on montre que l'on a un modèle décrit par la relation non linéaire suivante.

$$y_{2,i} = E\left(y_{2,i}^* / \, y_{1,i}^* > 0\right) + v_i$$

avec $v_i = y_{2,i} - E\left(y_{2,i}^* / y_{1,i}^* > 0\right)$ ou encore :

$$y_{2,i} = x_{2,i}\beta_2 + \sigma_{12}\sigma_1^{-1}\lambda\left(x_{1,i}\theta_1\right) + \widetilde{\mu}_{2,i} \tag{2.21}$$

où les perturbations $\widetilde{\mu}_{2,i}$ vérifient

$$\widetilde{\mu}_{2,i} = \mu_{2,i} + v_i = \varepsilon_{2,i} - \left(\sigma_{12}\sigma_1^{-2}\right)\varepsilon_{1,i} + \left[y_{2,i} - E\left(y_{2,i}^* / y_{1,i}^* > 0\right)\right] \tag{2.22}$$

On obtient ainsi la proposition suivante.

Proposition 2.5. ***Le modèle Tobit généralisé de type II, pour les observations*** $y_{2,i} > 0$, ***peut être représenté par la relation non linéaire hétéroscédastique suivante :***

$$y_{2,i} = x_{2,i}\beta_2 + \sigma_{12}\sigma_1^{-1}\lambda\left(x_{1,i}\theta_1\right) + \widetilde{\mu}_{2,i} \tag{2.23}$$

avec $\theta_1 = \beta_1/\sigma_1$ ***et où les perturbations*** $\widetilde{\mu}_{2,i}$ ***vérifient*** $E\left(\widetilde{\mu}_{2,i}\right) = 0$ ***et :***

$$Var\left(\widetilde{\mu}_{2,i}\right) = \sigma_2^2 - \sigma_{12}\sigma_1^{-2}\left[x_{1,i}\theta_1\lambda\left(x_{1,i}\theta_1\right) + \lambda\left(x_{1,i}\theta_1\right)^2\right] \tag{2.24}$$

Dans le cas du modèle Tobit généralisé de type II, la méthode d'estimation d'Heckman, dite aussi méthode d'estimation en deux étapes, en comporte en fait trois.

- **Etape 1** : On commence par estimer le ratio $\theta_1 = \beta_1/\sigma_1$ en utilisant la partie dichotomique du modèle, c'est à dire en modélisant la probabilité d'obtenir une valeur $y_{1,i}^*$ positive.

$$z_{1,i} = \begin{cases} 1 & \text{si } y_{1,i}^* > 0 \\ 0 & \text{si } y_{1,i}^* \leq 0 \end{cases} \qquad y_{1,i}^* \rightsquigarrow N\left(x_1\beta_1, \sigma_1^2\right) \tag{2.25}$$

  Pour cela, on considère **le modèle probit dichotmique** suivant :

$$Prob\left(z_{i,1} = 1\right) = Prob\left(y_{1,i}^* > 0\right) = \Phi\left(x_1\theta_1\right) \tag{2.26}$$

  Soit $\widehat{\theta}_1$ un estimateur convergent de θ_1 obtenu à partir de ce modèle probit.

- **Etape 2** : A partir de l'estimateur $\widehat{\theta}_1$ on constuit le ratio de Mill $\lambda\left(x_{1,i}\widehat{\theta}_1\right)$ pour chaque observation $x_{1,i}$. Soit $\widehat{\lambda}\left(x_{1,i}\widehat{\theta}_1\right)$ l'estimateur ainsi obtenu. On effectue alors la régression linéaire suivante par la méthode des MCO :

$$y_{2,i} = x_{2,i}\widehat{\beta}_2 + \hat{\sigma}\,\widehat{\lambda}\left(x_{1,i}\widehat{\theta}_1\right) + \widetilde{\mu}_{2,i} \tag{2.27}$$

  et l'**on obtient alors un estimateur asymptotiquement convergent des paramètres** β_2, **noté** $\widehat{\beta}_2$ **et un estimateur asymptotiquement convergent** $\hat{\sigma}$ **du ratio de paramètres** $\sigma_{12}\sigma_1^{-1}$. Si l'on impose une contrainte sur σ_1 (par exemple $\sigma_1 = 1$) cela permet alors d'identifier la covariance σ_{12}. Toutefois, ces deux estimateurs ne sont efficaces en rtaison de de l'hétéroscédascticité :

$$Var\left(\widetilde{\mu}_{2,i}\right) = \sigma_2^2 - \widehat{\sigma}\,\sigma_1^{-1}\left[x_{1,i}\widehat{\theta}_1\,\widehat{\lambda}\left(x_{1,i}\widehat{\theta}_1\right) + \widehat{\lambda}\left(x_{1,i}\widehat{\theta}_1\right)^2\right]$$

- **Etape 3** : Reste alors à estimer le paramètre σ_2. Pour cela, considérons le résidu $\widetilde{\mu}_{2,i}$ de la régression (2.27). Pour une valeur donnée de σ_1, on obtient alors par construction **un estimateur convergent de** σ_2^2 **:**

$$\widehat{\sigma}_2^2 = \frac{1}{N_1} \sum_{i:\, y_{2,i}>0} \widetilde{\mu}_{2,i} + \frac{\widehat{\sigma}}{\sigma_1} \frac{1}{N_1} \sum_{i:\, y_{2,i}>0} \left[x_{1,i}\widehat{\theta}_1 \, \widehat{\lambda} \left(x_{1,i}\widehat{\theta}_1 \right) + \widehat{\lambda} \left(x_{1,i}\widehat{\theta}_1 \right)^2 \right] \tag{2.28}$$

  où N_1 désigne le nombre d'observations pour lesquelles $y_{2,i} > 0$.

On montre que les estimateurs $\widehat{\beta}_2$, $\widehat{\sigma}_2^2$ et $\widehat{\sigma}_{12}$ ainsi obtenus sont asymptotiquement convergent et normalement distribuées (Olsen 1980).

Proposition 2.6. ***La convergence des estimateurs*** $\widehat{\beta}_2$, $\widehat{\sigma}_2^2$ ***et*** $\widehat{\sigma}_{12}$ ***de Heckman ne nécessite pas de supposer la normalité jointe des variables*** $y_{1,i}^*$ ***et*** $y_{2,i}^*$***. Il suffit de supposer que la variable*** $y_{1,i}^*$ ***est distribuée selon une loi normale et que la composante des résidus*** $\mu_{2,i}$ ***telle que***

$$y_{2,i}^* = x_{2,i}\beta_2 + \varepsilon_{2,i} = x_{2,i}\beta_2 + \sigma_{12}\sigma_1^{-2} \left(y_{1,i}^* - x_{1,i}\beta_1 \right) + \mu_{2,i}$$

est indépendamment distribuée par rapport à $y_{1,i}^*$***.***

Ainsi, contrairement à l'estimateur du *MV*, la convergence de l'estimateur de Heckman ne recquiert pas la normalité jointe des deux variables latentes.

2.1.4. Exemples

Un des exemples les plus célébres de modèle Tobit généralisés de type II est l'application de Gronau (1974) sur le travail des femmes. Son modèle d'offre de travail est fondé sur la théorie du salaire de réservation et sera repris par la suite de très nombreuses fois, notamment par Heckman avec des modèles de biais de sélection.

Le travail de Gronau (1974) consiste à déterminer sous quelles conditions les femmes décident de travailler ou de ne pas travailler. Gronau suppose que le taux de salaire réel effectivement offert aux femmes, noté W^s, est indépendant du nombre d'heures travaillées H. La femme maximise son utilité $U(C, X)$ où C désigne le temps passé à s'occuper des enfants et X le vecteur des autres biens de consommation. Etant donné le salaire W^s offert, la femme cherche donc le panier (C, X) qui maximise son utilité sous la contrainte de temps $C + H = T$ où T désigne temps disponible total et sous la contrainte de revenu $X = W^s H + V$ où V désigne le montant de ses revenus autres que les revenus salariaux. Le programme est donc :

$$\begin{aligned} &\max_{\{C,X\}} U(C, X) \\ &sc : C + H = T \\ &sc : X = W^s H + V \end{aligned} \tag{2.29}$$

Dès lors la femme n'acceptera de travailler que si le TMS du bien C au bien X évalué au point $H = 0$ (c'est à dire sans travailler) est inférieur le taux de salaire W^s.

$$\left. \frac{\partial U(C, X) / \partial C}{\partial U(C, X) / \partial X} \right|_{H=0} < W^s$$

Intuitivement, sans travailler, la réduction d'une unité du temps consacré aux enfants implique pour maintenir le niveau d'utilité une augmentation de la consommation égale au terme de gauche. Si en travaillant une unité la femme, gagne W^s et que ce salaire réel est supérieur à cette augmentation nécessaire de la consommation, la femme décidera de travailler. Si elle décide travailler, le nombre d'heures travaillées H sera défini par l'égalité :

$$\frac{\partial U(C,X)/\partial C}{\partial U(C,X)/\partial X}(H) = W^s$$

Gronau qualifie le terme de gauche de "housewife's value of time", mais plus généralemnt il s'agit ici d'un salaire de réservation, noté W^r. Si $W^r > W^s$, l'agent accepte de travailler, sinon il refuse.

En supposant que W^r et W^s peuvent s'écrire comme la somme de combinaisons linéaires des variables explicatives indépendantes et d'un terme d'erreur, le modèle devient :

$$W_i^s = x_{1,i}\beta_1 + \varepsilon_{1,i} \tag{2.30}$$

$$W_i^r = x_{2,i}\beta_2 + \varepsilon_{2,i} \tag{2.31}$$

$$W_i = \begin{cases} W_i^s & \text{si } W_i^s > W_i^r \\ 0 & \text{sinon} \end{cases} \tag{2.32}$$

où W_i désigne le salaire effectif. Pour les femmes ayant un salaire de réservation supérieur au salaire offert, le salaire effectif est nul puisque ces dernières refusent de travailler. En posant que $y_{1,i}^* = W_i^s - W_i^r$ et $y_{1,i}^* = W_i^s$, en supposant que les perturbations $\varepsilon_{1,i}$ et $\varepsilon_{2,i}$ sont normallement dsitribuées on retrouve un modèle tobit généralisé de type II. C'est un tel modèle qui donna lieu plus tard à l'extension des modèles à troncature auxiliaire ou modèles Heckit.

2.1.5. Modèle de Troncature Auxiliaire ou Modèle Heckit

Parfois, on appele **modèle de Troncature Auxiliaire** ou **modèle Heckit**, en hommage à Heckman, ou **modèles de biais de sélection**, les modèles Tobit généralisés de type II. Les exemples sont multiples : caractéristiques des demandeurs d'emploi connues que s'ils sont inscrits à l'ANPE, notes des étudiants connues que s'ils ont décidé de passer l'examen, réponses aux enquêtes connues que si les individus ont décidé de les fournir, etc... **Tous ces exmples cachent un processus de selction des individus observés dans lequel ceux-ci interviennent de façon déterminante : on a donc un problème d'auto-sélection**.

Considérons un modèle Tobit généralisé :

$$y_{2,i} = \begin{cases} y_{2,i}^* & \text{si } y_{1,i}^* > 0 \\ 0 & \text{si } y_{1,i}^* \leq 0 \end{cases} \tag{2.33}$$

$$y_{1,i}^* = x_{1,i}\beta_1 + \varepsilon_{1,i} \tag{2.34}$$

$$y_{2,i}^* = x_{2,i}\beta_2 + \varepsilon_{2,i} \tag{2.35}$$

Il est tentant alors d'appliquer les MCO à l'ensemble des observations pour déterminer les paramètres β_2.

Proposition 2.7. ***On montre que l'estimateur $\widehat{\beta}_2$ des MCO ne sera pas biaisé qu'à partir du moment où le processus de sélection sera totalement indépendant de la variable auxiliaire $y_{1,i}^*$:***

$$E\left(\widehat{\beta}_2\right) = \beta_2 \Longleftrightarrow \rho\sigma_1\lambda\left(x_{1,i}\beta_1\right) = 0 \Longleftrightarrow \rho = 0 \tag{2.36}$$

Dès que le processus de sélevction dépend, même partiellement de $y_{1,i}^*$, le biais s'introduit : il s'agit d'**un biais de sélection**. Le biais apparaît parce que certaines variables explicatives, celles contenues dans $x_{1,i}$, ont été oubliées : il s'agit encore une fois d'un biais de variable omise. On parle parfois de modèle Heckit, en hommage à Heckman, la spécification de ce dernier étant donnée par :

$$z_i^* = w_i\gamma + \mu_i \tag{2.37}$$

$$y_i = x_i\beta + \varepsilon_i \tag{2.38}$$

où y_i n'est observée que si $z_i^* > 0$ ou encore si $z_i = 1$ avec :

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

2.2. Autres Modèles Tobit Généralisés

2.2.1. Modèle Tobit Généralisé Type 3

2.2.2. Modèle Tobit Généralisé Type 4

2.2.3. Modèle Tobit Généralisé Type 5

3. Les Modèles à régimes

3.1. Modèle à régimes observables

3.2. Modèle à régimes inobservables

A. Annexes

A.1. Concavité de la log-vraisemblance

Soit les matrices Δ et Γ respectivement définies par :

$$\Delta = \begin{pmatrix} \Psi(\alpha, h) & 0 \\ 0 & -\frac{N_1}{h^2} \end{pmatrix} \tag{A.1}$$

$$\Gamma = \begin{pmatrix} -\sum_{i:\, y_i>0} x_i x_i' & \sum_{i:\, y_i>0} x_i y_i \\ \sum_{i:\, y_i>0} y_i x_i' & \sum_{i:\, y_i>0} y_i^2 \end{pmatrix} \tag{A.2}$$

avec

$$\Psi(\alpha, h) = \sum_{i:\, y_i=0} \frac{\phi(x_i\alpha)}{1-\Phi(x_i\alpha)} \left[x_i\alpha - \frac{\phi(x_i\alpha)}{1-\Phi(x_i\alpha)} \right] x_i x_i'$$

et $x_i\alpha - \phi(x_i\alpha)\left[1-\Phi(x_i\alpha)\right]^{-1} < 0$. Sachant que la matrice $x_i x_i'$ est définie positive.

*** A finir ***

$\begin{pmatrix} a & b \\ b & c \end{pmatrix}$, eigenvalues:

$\frac{1}{2}a + \frac{1}{2}c + \frac{1}{2}\sqrt{(a^2 - 2ac + c^2 + 4b^2)}, \frac{1}{2}a + \frac{1}{2}c - \frac{1}{2}\sqrt{(a^2 - 2ac + c^2 + 4b^2)}$

A.2. Programme de simulation d'un probit simple

Le programme permettant de simuler la série observable y_i est le suivant :

```
' Tirage des Epsilon dans une loi N(0,1)
scalar sigeps=1
genr eps=nrnd*sigeps
' Construction de la Variable Exogène X
scalar sigx=1
genr x=nrnd*sigx
' Construction de la Variable Latente y*
scalar beta=0.8
scalar alpha=1
genr ystar=alpha+beta*x+eps
' Construction de la Variable Observale y
genr y=0
genr y= (ystar>0)*ystar
```

Bibliographie

Alban T. (2000), "Econométrie des Variables Qualitatives", *Dunod.*

Goldberger (1964)
Gourieroux C. (1989), "Econométrie des Variables Qualitatives", *Economica.*

Greene W.H. (1997), "Econometric Analysis", *Londres, Prentice Hall.*

McDonald, J. and R. Moffitt (1980) The Uses of Tobit Analysis, Review of Economic and Statistics, 62, 318321

Maddala. G.S. (1983), "Limited-dependent and Qualitative Variables in Econometrics", *Econometric Society Monographs,* 3, Cambrige University Press.

Tobin J. (1958), "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, 26, 24-36.