ÉCONOMÉTRIE
(UGA S2)

Modèles de choix dichotomique : Probit et Logit

Michal W. Urdanivia*

*Université de Grenoble Alpes, Faculté d'Économie, GAEL,
e-mail: michal.wong-urdanivia@univ-grenoble-alpes.fr

3 avril 2024

1. Structure des modèles économétriques de choix dichotomique

- L'individu *i* a le choix entre deux options, 1 ou 0. Ce choix est décrit par :

$$\begin{vmatrix} y_i = 1 \ \text{ si } \ i \text{ choisit l'option 1} \\ y_i = 0 \ \text{ si } \ i \text{ choisit l'option 0} \end{vmatrix}$$

- La variable latente y_i^* mesure le ***bénéfice net du choix de l'option*** 1 par rapport à celui de l'option 0, *e.g.* $U_{1,i}^* - U_{0,i}^*$.

- Si *i* est ***rationnel*** on a alors :

$$\boxed{\begin{vmatrix} y_i = 1 \ \text{ si } \ y_i^* > 0 \\ y_i = 0 \ \text{ sinon} \end{vmatrix} \qquad \textbf{\textit{ou}} \qquad y_i = \mathbf{1}(y_i^* > 0)}$$

- Ce modèle de choix dichotomique est parfois qualifié de « ***modèle à index*** (unique) », la variable latente y_i^* étant l' « index »

L’individu *i* est décrit par $\mathbf{x}_i$ et on pose :

$$\boxed{y_i^* = \mathbf{x}_i'\mathbf{a}_0 + u_i \text{ avec } E[u_i] \equiv 0 \text{ et, } u_i \text{ et } \mathbf{x}_i \text{ indépendantes}}$$

- Le terme d’erreur u_i contient les effets sur y_i^* de ce qui n’est pas mesuré par $\mathbf{x}_i'\mathbf{a}_0$. Il s’agit des effets de :
  - ***Hétérogénéité*** (« permanente ») des ***préférences/besoins individuels*** expliquant le bénéfice net de choisir 1 plutôt que 0.
  - ***Emotions*** (« transitoires ») affectant le choix décrit par y_i, au moment de ce choix.

- Le terme d’erreur u_i porte le caractère probabiliste du modèle.

2. Modèles Logit et Probit

- La forme de $P_{y_i/\mathbf{x}_i}\left[y_i=1\right]=1-P_{y_i/\mathbf{x}_i}\left[y_i=0\right]$ dépend de la distribution conditionnelle du terme d'erreur u_i, *i.e.* de $u_i/\mathbf{x}_i$.

- Deux choix classiques pour $u_i/\mathbf{x}_i$:

$u_i/\mathbf{x}_i \sim \mathcal{N}(0;1)$, loi normale centrée réduite : ***modèle Probit*** (***Normit***)

$u_i/\mathbf{x}_i \sim \mathcal{L}(0,1)$, loi logistique standard : ***modèle Logit***

Rmq. Ces choix ***définissent entièrement*** la loi de $u_i/\mathbf{x}_i$, *e.g.* dans le cas du modèle Probit $V\left[u_i/\mathbf{x}_i\right]=V\left[u_i\right]=1$. Il s'agit d'un choix de ***normalisation***.

Définition et propriétés. ***Loi Logistique***

La variable aléatoire $\tilde{x}$ suit une loi logistique de paramètres μ (localisation) et s (échelle, $s > 0$), notée ici $\mathcal{L}(\mu, s)$, si :

$$F_{\tilde{x}}(x) = \left[1 + \exp\left(-s^{-1} \times (x - \mu)\right)\right]^{-1} \text{ pour } x \in \mathbb{R}$$

Fonction de densité :

$$f_{\tilde{x}}(x) = \exp\left(-s^{-1} \times (x - \mu)\right) \times s^{-1} \times \left[1 + \exp\left(-s^{-1} \times (x - \mu)\right)\right]^{-2}$$

Moments :

$$E[\tilde{x}] = \mu \text{ et } V[\tilde{x}] = s^2 \times \pi^2 / 3$$

La logistique est symétrique est unimodale et a la forme « en cloche » de la loi normale mais avec des queues de distribution plus épaisses.

Loi logistique standard : $\mu = 0$ et $s = 1$

$$\tilde{x}_0 \sim \mathcal{L}(0,1) \text{ et } \tilde{x}_0 = \mu + s \times \tilde{x} \quad \Rightarrow \quad \tilde{x} \sim \mathcal{L}(\mu, s)$$

Calcul des fonctions de probabilité des choix

- ***Dans tous les cas on a :***

$$\boxed{P_{y_i/\mathbf{x}_i}\left[y_i=1\right]=1-F_{u_i/\mathbf{x}_i}(-\mathbf{x}_i'\mathbf{a}_0) \text{ et } P_{y_i/\mathbf{x}_i}\left[y_i=0\right]=F_{u_i/\mathbf{x}_i}(-\mathbf{x}_i'\mathbf{a}_0)}$$

car avec :

$$y_i=1 \quad \Leftrightarrow \quad y_i^*=\mathbf{x}_i'\mathbf{a}_0+u_i>0$$

on a :

$$P_{y_i/\mathbf{x}_i}\left[y_i=1\right]=P_{y_i^*/\mathbf{x}_i}\left[y_i^*>0\right]=P_{u_i/\mathbf{x}_i}\left[u_i>-\mathbf{x}_i'\mathbf{a}_0\right]=1-F_{u_i/\mathbf{x}_i}(-\mathbf{x}_i'\mathbf{a}_0)$$

et :

$$P_{y_i/\mathbf{x}_i}\left[y_i=0\right]=1-P_{y_i/\mathbf{x}_i}\left[y_i=1\right]=F_{u_i/\mathbf{x}_i}(-\mathbf{x}_i'\mathbf{a}_0).$$

▪ ***Cas du modèle Probit*** :

$$\left| \begin{array}{l} y_i = 1 \ \text{ si } y_i^* > 0 \\ y_i = 0 \ \text{ si } y_i^* \leq 0 \end{array} \right. \quad \text{avec} \quad y_i^* \big/ \mathbf{x}_i \sim \mathcal{N}(\mathbf{x}_i'\mathbf{a}_0 ; 1)$$

On a :

$$\boxed{P_{y_i/\mathbf{x}_i}\left[y_i = 1\right] = \Phi(\mathbf{x}_i'\mathbf{a}_0) \quad \text{et} \quad P_{y_i/\mathbf{x}_i}\left[y_i = 0\right] = 1 - \Phi(\mathbf{x}_i'\mathbf{a}_0)}$$

car :

$$\begin{aligned} P_{y_i/\mathbf{x}_i}\left[y_i = 1\right] = P_{y_i^*/\mathbf{x}_i}[y_i^* > 0] &= P_{u_i/\mathbf{x}_i}\left[\mathbf{x}_i'\mathbf{a}_0 + u_i > 0\right] \\ &= P_{u_i/\mathbf{x}_i}\left[u_i > -\mathbf{x}_i'\mathbf{a}_0\right] \\ &= 1 - \Phi(-\mathbf{x}_i'\mathbf{a}_0) = \Phi(\mathbf{x}_i'\mathbf{a}_0) \end{aligned}$$

et :

$$P_{y_i/\mathbf{x}_i}\left[y_i = 0\right] = 1 - P_{y_i/\mathbf{x}_i}\left[y_i = 1\right] = 1 - \Phi(\mathbf{x}_i'\mathbf{a}_0).$$

- ***Cas du modèle Logit*** :

$$\begin{vmatrix} y_i = 1 \ \text{ si } y_i^* > 0 \\ y_i = 0 \ \text{ si } y_i^* \le 0 \end{vmatrix} \quad \text{avec} \quad y_i^* \mid \mathbf{x}_i \sim \mathcal{L}(\mathbf{x}_i'\mathbf{a}_0, 1)$$

On a :

$$\boxed{P_{y_i/\mathbf{x}_i}\left[y_i = 1\right] = \frac{\exp(\mathbf{x}_i'\mathbf{a}_0)}{1+\exp(\mathbf{x}_i'\mathbf{a}_0)} \text{ et } P_{y_i/\mathbf{x}_i}\left[y_i = 0\right] = \frac{1}{1+\exp(\mathbf{x}_i'\mathbf{a}_0)}.}$$

Rmq. Les fonctions de probabilité obtenues à partir des modèles ***Logit*** et ***Probit*** sont proches.

$$\textbf{\textit{Probit :}}\ E\left[u_i\right]=0 \text{ et } V\left[u_i\right]=1$$

$$\textbf{\textit{Logit :}}\ E\left[u_i\right]=0 \text{ et } V\left[u_i\right]=\pi^2/3$$

$$\Phi(\mathbf{x}_i'\mathbf{a}) \simeq \frac{\exp(\mathbf{x}_i'\boldsymbol{\alpha})}{1+\exp(\mathbf{x}_i'\boldsymbol{\alpha})} \quad \text{avec } \boldsymbol{\alpha}=\frac{\sqrt{3}}{\pi}\times\mathbf{a}$$

Rmq. La fonction de répartition de la loi logistique a des ***queues de distribution*** plus épaisses que celles de la loi normale centrée réduite.

Rmq. La loi Logistique est également une différence de lois à valeurs extrêmes de type I (EV I). La loi EV I est max-stable et est donc appropriée pour représenter des niveaux d'utilité « optimisés »

3. Identification et estimation des paramètres des modèles Probit et Logit

- On suppose que les $(y_i, \mathbf{x}_i)$ sont iid pour $i = 1, ..., N$ et on veut calculer :

$$\hat{\mathbf{a}}_N^{MV} \equiv \arg\max_{\mathbf{a}} L_N(\mathbf{a})$$

où :

$L_N(\mathbf{a})$ est la log-vraisemblance en **a** des observations de l'échantillon.

- Les $(y_i, \mathbf{x}_i)$ sont iid pour $i = 1, ..., N$ on a donc :

$$L_N(\mathbf{a}) \equiv \ln \prod_{i=1}^{N} \ell_i(\mathbf{a}) = \sum_{i=1}^{N} \ln \ell_i(\mathbf{a})$$

- La ***forme de la vraisemblance*** en **a** de l'observation i, $\ell_i(\mathbf{a})$, dépend ***de la valeur de*** y_i.

- Deux cas sont à distinguer ici :

$$\ell_i(\mathbf{a}) = \begin{cases} 1 - F_{u_i/\mathbf{x}_i}(-\mathbf{x}_i'\mathbf{a}) & \text{si } y_i = 1 \\ F_{u_i/\mathbf{x}_i}(-\mathbf{x}_i'\mathbf{a}) & \text{si } y_i = 0 \end{cases}$$

- La ***forme de la vraisemblance*** en **a** de l'observation i, $\ell_i(\mathbf{a})$, dépend ***de la valeur de*** y_i. On peut alors définir $\ell_i(\mathbf{a})$ par :

$$\ell_i(\mathbf{a}) = \left[1 - F_{u_i/\mathbf{x}_i}(-\mathbf{x}_i'\mathbf{a})\right]^{y_i} \times \left[F_{u_i/\mathbf{x}_i}(-\mathbf{x}_i'\mathbf{a})\right]^{1-y_i}$$

et $\ln \ell_i(\mathbf{a})$ par :

$$\begin{aligned} \ln \ell_i(\mathbf{a}) &= y_i \times \ln\left[1 - F_{u_i/\mathbf{x}_i}(-\mathbf{x}_i'\mathbf{a})\right] + (1 - y_i) \times \ln F_{u_i/\mathbf{x}_i}(-\mathbf{x}_i'\mathbf{a}) \\ &= y_i \times \ln \Phi(\mathbf{x}_i'\mathbf{a}) + (1 - y_i) \times \ln\left[1 - \Phi(\mathbf{x}_i'\mathbf{a})\right] \qquad (\textbf{\textit{probit}}) \\ &= y_i \times \ln\left[\frac{\exp(\mathbf{x}_i'\mathbf{a})}{1 + \exp(\mathbf{x}_i'\mathbf{a})}\right] + (1 - y_i) \times \ln\left[\frac{1}{1 + \exp(\mathbf{x}_i'\mathbf{a})}\right] \qquad (\textbf{\textit{logit}}) \end{aligned}$$

- Log-vraisemblance de l'échantillon :

Probit
$L_N(\mathbf{a}) = \sum_{i=1}^{N} \ln \ell_i(\mathbf{a}) = \sum_{i=1}^{N} \left\{ y_i \times \ln \Phi(\mathbf{x}_i'\mathbf{a}) + (1-y_i) \times \ln\left[1-\Phi(\mathbf{x}_i'\mathbf{a})\right] \right\}$

Logit
$L_N(\mathbf{a}) = \sum_{i=1}^{N} \left\{ y_i \times \ln\left[\frac{\exp(\mathbf{x}_i'\mathbf{a})}{1+\exp(\mathbf{x}_i'\mathbf{a})}\right] + (1-y_i) \times \ln\left[\frac{1}{1+\exp(\mathbf{x}_i'\mathbf{a})}\right] \right\}$

- Dans les deux cas, $E\left[\ln \ell_i(\mathbf{a})\right]$ est ***strictement concave*** en **a** si et seulement si $E\left[\mathbf{x}_i\mathbf{x}_i'\right]$ est inversible.
    - Si $\mathbf{x}_i$ ne contient pas un prédicteur « parfait » de y_i

Le paramètre $\mathbf{a}_0$ est donc identifiable si et seulement si $E\left[\mathbf{x}_i\mathbf{x}_i'\right]$ est inversible.

- Les CO1 de la maximisation de la log-vraisemblance sont données par :

$$N^{-1}\sum_{i=1}^{N}\frac{f_{u_i|\mathbf{x}_i}(\mathbf{x}_i'\hat{\mathbf{a}}_N^{MV})\left(y_i - F_{u_i|\mathbf{x}_i}(\mathbf{x}_i'\hat{\mathbf{a}}_N^{MV})\right)}{F_{u_i|\mathbf{x}_i}(\mathbf{x}_i'\hat{\mathbf{a}}_N^{MV})\left(1-F_{u_i|\mathbf{x}_i}(\mathbf{x}_i'\hat{\mathbf{a}}_N^{MV})\right)}\mathbf{x}_i = \mathbf{0}$$

et la distribution *as*. de $\hat{\mathbf{a}}_N^{MV}$ est donnée par :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MV} - \mathbf{a}_0)\xrightarrow[N\to+\infty]{L}\mathcal{N}\left(\mathbf{0}, E\left[\frac{f_{u_i|\mathbf{x}_i}(\mathbf{x}_i'\mathbf{a}_0)^2}{F_{u_i|\mathbf{x}_i}(\mathbf{x}_i'\mathbf{a}_0)\left(1-F_{u_i|\mathbf{x}_i}(\mathbf{x}_i'\mathbf{a}_0)\right)}\mathbf{x}_i\mathbf{x}_i'\right]\right).$$

Remarque importante quant à l'identification des paramètres des modèles économétriques paramétriques de choix dichotomique

- On utilise ici l'exemple d'un modèle Probit, sachant que ce qui suit est valide pour tout modèle économétrique de choix discret.
- On aurait pu spécifier le modèle suivant :

$$\left| \begin{array}{l} y_i = 1 \ \text{ si } y_i^* > 0 \\ y_i = 0 \ \text{ si } y_i^* \le 0 \end{array} \right. \quad \text{avec} \quad y_i^* \big/ \mathbf{x}_i \sim \mathcal{N}(\mathbf{x}_i' \mathbf{a}_0 ; \sigma_0^2),$$

  i.e., ne pas imposer $\sigma_0^2 = 1$.

- La log-vraisemblance de $y_i \big/ x_i$ en $(\mathbf{a}, \omega)$ s'écrit :

$$\ln \ell_i(\mathbf{a}) = y_i \times \ln \Phi(\mathbf{x}_i' \mathbf{a} \sigma^{-1}) + (1 - y_i) \times \ln \left[1 - \Phi(\mathbf{x}_i' \mathbf{a} \sigma^{-1}) \right]$$

- Dans cette log-vraisemblance, les paramètres $\mathbf{a}$ et σ apparaissent toujours « ensemble », *i.e.* sous la forme $\mathbf{a}\sigma^{-1}$.

- Ceci implique que dans le modèle :

$$\left|\begin{array}{l} y_i = 1 \ \text{ si } y_i^* > 0 \\ y_i = 0 \ \text{ si } y_i^* \leq 0 \end{array}\right. \quad \text{avec} \quad y_i^* \big/ \mathbf{x}_i \sim \mathcal{N}(\alpha_0 + \tilde{\mathbf{x}}_i' \mathbf{b}_0 ; \sigma_0^2),$$

  les paramètres $\mathbf{a}_0$ ***et*** σ_0 ***ne sont pas identifiables séparément***.

- σ_0 est un paramètre de dispersion (donc d'échelle) de y_i^* dont on n'observe que le signe. C'est ce qui impose des ***choix dits de normalisation*** :
  - Jusqu'à présent on a utilisé $\sigma_0 \equiv 1$ …
    - « Fixe » l'échelle de mesure des utilités lorsque $y_i^* \equiv U_{1,i}^* - U_{0,i}^*$
  - … mais on utilise parfois $b_{r,0} \equiv 1$ ou $b_{r,0} \equiv -1$ pour $r \in \{1, ..., K\}$
    - Utilisé, *e.g.*, lorsque y_i^* mesure un « bénéfice net » exprimé en unités monétaires et lorsque x_{ri} est un élément du calcul de ce bénéfice, *i.e.* une partie du coût ou du bénéfice brut)

Lorsqu'on choisit $\sigma_0 \equiv 1,\ y_i^* = \alpha_0 + \mathbf{b}_0'\mathbf{x}_i + u_i$ et $V\left[u_i\right] \equiv 1$

- Les ***valeurs absolues*** de α_0 ou des éléments de $\mathbf{b}_0$ n'ont ***pas d'interprétation*** …
- … mais :
    - ***les signes des paramètres*** $b_{k,0}$ peuvent être interprétés
    - ***les valeurs des paramètres*** $b_{k,0}$ ***et*** $b_{\ell,0}$ ***peuvent être comparées*** si $x_{k,i}$ et $x_{\ell,i}$ sont mesurées dans la ***même unité***.
- Des $\mathbf{b}_0$ estimés avec de petites valeurs absolues est « mauvais signe » :

$$V[y_i^*] = V\left[\mathbf{x}_i'\mathbf{b}_0\right] + V\left[u_i\right] = \mathbf{b}_0'V\left[\mathbf{x}_i\right]\mathbf{b}_0 + 1,$$

la variance de la partie expliquée de y_i^* est petite par rapport à sa variance non expliquée qui est 1 ici, par construction.

Lorsqu'on choisit $b_{1,0} \equiv \pm 1$, $y_i^* = \alpha_0 \pm b_{1,0} x_{1,i} + \mathbf{b}'_{-1,0}\mathbf{x}_{-1,i} + u_i$ et $V[u_i] \equiv 1$

- Dans le cas $b_{1,0} \equiv 1$, on a $y_i^* = \alpha_0 + x_{1,i} + \mathbf{x}'_{-1,i}\mathbf{b}_{-1,0} + u_i$ avec $u_i / \mathbf{x}_i \sim \mathcal{N}(0;\sigma_0^2)$ et :

$$\ln \ell_i(\alpha, \mathbf{b}_{-1}, \sigma) = y_i \times \ln \Phi\left[\sigma^{-1}(\alpha + x_{1,i} + \mathbf{x}'_{-1,i}\mathbf{b}_{-1})\right]$$
$$+ (1 - y_i) \times \ln\left[1 - \Phi\left[\sigma^{-1}(\alpha + x_{1,i} + \mathbf{x}'_{-1,i}\mathbf{b}_{-1})\right]\right].$$

- Les paramètres α_0, $\mathbf{b}_{-1,0}$ et σ_0 sont alors identifiables séparément (les variations de $x_{1,i}$ identifiant σ_0 dès lors que $E[\mathbf{x}_i\mathbf{x}'_i]$ est inversible).

- La ***qualité prédictive du modèle*** est médiocre (bonne) lorsque l'estimation de σ_0^2 est élevée (petite) par rapport à la variance de $\alpha_0 + x_{1,i} + \mathbf{x}'_{-1,i}\mathbf{b}_{-1,0}$.

- Compte-tenu de ce que :

$$P\left[y_i / \mathbf{x}_i\right] = \Phi\left[\sigma_0^{-1}(\alpha_0 + x_{1,i} + \mathbf{x}'_{-1,i}\mathbf{b}_{-1,0})\right]$$

on a :

$$\lim_{\sigma_0 \to +\infty} \underbrace{P\left[y_i / \mathbf{x}_i\right] = 1/2}_{\text{Choix parfaitement aléatoire}}$$

et :

$$\lim_{\sigma_0 \to 0} \underbrace{P\left[y_i / \mathbf{x}_i\right] = \mathbf{1}\left[\alpha_0 + x_{1,i} + \mathbf{x}'_{-1,i}\mathbf{b}_{-1,0} > 0\right]}_{\text{Choix parfaitement expliqué par } \mathbf{x}_i}.$$

4. Prédiction et mesures d'ajustement

- Il existe plusieurs mesures d'ajustement des estimations des modèles économétriques de choix dichotomiques : le ρ^2 ***de McFadden*** et des mesures d'ajustement fondées sur des ***prédictions de*** y_i.

- Le ρ^2 ***de McFadden*** vise à mesurer l'apport des variables explicatives du modèles, comme le R^2 d'un modèle de régression linéaire. Il utilise le fait qu'on a généralement (avec $\sigma_0 \equiv 1$ dans le cas du Probit) :

$$y_i^* = \alpha_0 + \mathbf{b}_0' \tilde{\mathbf{x}}_i + u_i$$

- La log-vraisemblance en la valeur estimée des paramètres est donnée par :

$$L_N(\hat{\mathbf{a}}_N^{MV}) = L_N(\hat{\alpha}_N^{MV}, \hat{\mathbf{b}}_N^{MV}) \leq 0.$$

  Pour mesurer l'apport des $\tilde{\mathbf{x}}_i$, on estime le modèle en imposant $\mathbf{b}_0 = \mathbf{0}$, i.e. sans l'apport d'information des $\tilde{\mathbf{x}}_i$ en comparant la log-vraisemblance obtenue avec $L_N(\hat{\alpha}_N^{MV}, \hat{\mathbf{b}}_N^{MV})$.

- Le modèle considéré lorsque $\mathbf{b}_0 = \mathbf{0}$ dit essentiellement que :

$$P[y_i = 1] = \pi_0 \quad \text{avec } \pi_0 = \begin{vmatrix} \Phi[\alpha_0] & \text{(Probit)} \\ \exp(\alpha_0)[1+\exp(\alpha_0)]^{-1} & \text{(Logit)} \end{vmatrix}$$

Il est alors aisé de montrer que :

$$\hat{\pi}_N^{MV} = N^{-1}\sum_{i=1}^{N} y_i = \begin{vmatrix} \Phi\left[\hat{\alpha}_N^{MVC}\right] & \text{(Probit)} \\ \exp(\hat{\alpha}_N^{MVC})\left[1+\exp(\hat{\alpha}_N^{MVC})\right]^{-1} & \text{(Logit)} \end{vmatrix}$$

On a alors :

$$L_N(\hat{\alpha}_N^{MVC}, \mathbf{0}) = N_1 \ln \hat{\pi}_N^{MV} + N_0 \ln(1 - \hat{\pi}_N^{MV})$$

avec :

$$N_1 \equiv \sum_{i=1}^{N} y_i \text{ et } N_0 \equiv \sum_{i=1}^{N} (1 - y_i), \text{ et donc } \hat{\pi}_N^{MV} = N_1 / N \text{ et } 1 - \hat{\pi}_N^{MV} = N_0 / N$$

.

- Le ρ^2 ***de McFadden*** est donné par :

$$\boxed{\rho^2 \equiv 1 - \frac{L_N(\hat{\alpha}_N^{MV}, \hat{\mathbf{b}}_N^{MV})}{L_N(\hat{\alpha}_N^{MVC}, \mathbf{0})} \in [0,1[\,.}$$

Il est compris entre 0 (inclus) et 1 (exclus) car :

$$L_N(\hat{\alpha}_N^{MVC}, \mathbf{0}) \leq L_N(\hat{\alpha}_N^{MV}, \hat{\mathbf{b}}_N^{MV}) < 0.$$

- On a généralement ρ^2 de l'ordre de 0.10, ce qui indique que les $\tilde{\mathbf{x}}_i$ accroissent de 10% la log-vraisemblance du modèle « simple »
  - De manière générale, les ***qualités prédictives des modèles économétriques de choix discret sont*** « ***modestes*** », indiquant que l'aléa de u_i ***joue un rôle important*** par rapport à l'information apportée par $\tilde{\mathbf{x}}_i$.
  - Les ***qualités prédictives des modèles Logit et Probit*** sont généralement ***proches de celles du modèle de probabilité linéaire***.

Prédire y_i et mesures de la qualité de prédiction

- La structure du modèle de y_i est donnée par :

$$\left| \begin{array}{l} y_i = 1 \text{ si } y_i^* = \mathbf{x}_i'\mathbf{a}_0 + u_i > 0 \\ y_i = 0 \text{ si } y_i^* = \mathbf{x}_i'\mathbf{a}_0 + u_i \leq 0 \end{array} \right. \quad \text{avec} \quad E\left[y_i^* / \mathbf{x}_i \right] = \mathbf{x}_i'\mathbf{a}_0$$

- Il semble donc naturel de ***prédire*** y_i ***par*** $\hat{y}_{i,N}$ tel que :

$$\left| \begin{array}{l} \hat{y}_{i,N} = 1 \text{ si } \hat{y}_{i,N}^* \equiv \mathbf{x}_i'\hat{\mathbf{a}}_N^{MV} > 0 \\ \hat{y}_{i,N} = 0 \text{ si } \hat{y}_{i,N}^* \equiv \mathbf{x}_i'\hat{\mathbf{a}}_N^{MV} \leq 0 \end{array} \right.$$

  puisque $\mathbf{x}_i'\hat{\mathbf{a}}_N^{MV}$ est un estimateur convergent de $E\left[y_i^* / \mathbf{x}_i \right] = \mathbf{x}_i'\mathbf{a}_0$.

- On peut alors construire différents ***indicateurs de la qualité de prédiction du modèle estimé*** à partir de $\hat{y}_{i,N}$, qui est un ***prédicteur de*** y_i construit à partir du modèle.

Indicateur de la qualité prédictive du modèle estimé

Part de prédictions correctes :

$$\hat{p} \equiv N^{-1}\sum_{i=1}^{N}\left[\hat{y}_{i,N}y_i+(1-\hat{y}_{i,N})(1-y_i)\right]$$

Part de prédictions correctes pour les $y_i=1$:

$$\hat{p}_{11} \equiv N_1^{-1}\sum_{i=1}^{N}\hat{y}_{i,N}y_i$$

Part de prédictions correctes pour les $y_i=0$:

$$\hat{p}_{00} \equiv N_0^{-1}\sum_{i=1}^{N}(1-\hat{y}_{i,N})(1-y_i)$$

Somme des parts de prédictions correctes pour les $y_i=1$ et les $y_i=0$:

$$\hat{p}_{11}+\hat{p}_{00}=N_1^{-1}\sum_{i=1}^{N}\hat{y}_{i,N}y_i+N_0^{-1}\sum_{i=1}^{N}(1-\hat{y}_{i,N})(1-y_i)$$

- Le pourcentage de prédictions correctes, $\hat{p}$, n'est pas un très bon indicateur :

$$\hat{p} = \frac{N_1}{N}\,\hat{p}_{11} + \frac{N_0}{N}\,\hat{p}_{00} = \hat{\pi}_N^{MV}\,\hat{p}_{11} + (1 - \hat{\pi}_N^{MV})\,\hat{p}_{00}\,.$$

  Il dépend des proportions de $y_i = 1$ et de $y_i = 0$, contrairement à $\hat{p}_{11}$ et $\hat{p}_{00}$.

- Il arrive souvent que l'un des indicateurs $\hat{p}_{11}$ et $\hat{p}_{00}$ est proche de 1 alors que l'autre est petit. *E.g.*, si $\hat{p}_{11} \simeq 1$ et $\hat{p}_{00} \simeq 0.25$ alors le modèle prédit correctement les $y_i = 1$ mais ne parvient pas à prédire « correctement » les $y_i = 0$.
    - Dans ce cas, le ***pouvoir de prédiction du modèle*** est « ***asymétrique*** ».

- Cela peut indiquer qu'il ***manque dans*** $\tilde{\mathbf{x}}_i$ ***des variables décisives pour expliquer le choix de l'option*** 0.

- En fait, il faut comparer les $\hat{p}_{11}$, $\hat{p}_{00}$ et $\hat{p}_{11}+\hat{p}_{00}$ par rapport à leurs équivalents dans le modèle le plus simple, *i.e.* avec $\mathbf{b}_0=\mathbf{0}$. Ce modèle donne les prédicteurs suivants :

$$\left|\begin{array}{l}\hat{y}_{i,N}^{C}=1 \text{ si } \alpha_N^{MVC}>0\\ \hat{y}_{i,N}^{C}=0 \text{ si } \alpha_N^{MVC}\leq 0\end{array}\right. \quad \text{ou} \quad \left|\begin{array}{l}\hat{y}_{i,N}^{C}=1 \textbf{ pour tout } i \text{ si } \hat{\pi}_N^{MV}>1/2\\ \hat{y}_{i,N}^{C}=0 \textbf{ pour tout } i \text{ si } \hat{\pi}_N^{MV}\leq 1/2\end{array}\right.$$

On a alors :

$$\left|\begin{array}{l}\hat{p}_{11}^{C}=\hat{\pi}_N^{MV} \text{ et } \hat{p}_{00}^{C}=1-\hat{\pi}_N^{MV} \text{ si } \hat{\pi}_N^{MV}>1/2\\ \hat{p}_{11}^{C}=1-\hat{\pi}_N^{MV} \text{ et } \hat{p}_{00}^{C}=\hat{\pi}_N^{MV} \text{ si } \hat{\pi}_N^{MV}\leq 1/2\end{array}\right.$$

- ***Rmq*** : si $\hat{\pi}_N^{MV}=N_1/N=0.90$ alors $\hat{p}_{11}^{C}=0.90$!

- De fait, le ***modèle estimé est considéré comme bon en prédiction si*** :

$$\boxed{\hat{p}_{11} + \hat{p}_{00} > 1}$$

  - En général, comme indiqué par l'ordre de grandeur habituel du ρ^2 de McFadden, la valeur de $\hat{p}_{11} + \hat{p}_{00}$ n'est que faiblement supérieure à 1.

Remarque générale

Ne pas trop se focaliser sur la prédiction quand-même.

Les modèles économétriques de choix discret sont par nature probabilistes …
… *i.e.* tiennent explicitement compte de ce que les choix considérés sont difficiles à prédire (!)

5. Exemples d'utilisation

Le scoring

- Une banque dispose de beaucoup de données, en particulier sur ses emprunteurs, … et elle s'en sert.
- Pour $i = 1, \ldots, N$ (N très grand), on a :

$$\left| \begin{array}{l} y_i = 1 \text{ si } i \text{ est un bon client} \\ y_i = 0 \text{ si } i \text{ a eu des problèmes de remboursement} \end{array} \right.$$

et :

$\mathbf{x}_i$: caractéristiques de i

(revenu, âge, ménage, situation financière initiale, emprunt, fonctionnaire, conjoncture économique, …, 1).

- Un modèle économétrique de choix dichotomique permet d’estimer la probabilité de défaut de *i* sous la forme :

$$\hat{P}_N\left[y_i = 0 / \mathbf{x}_i\right] = 1 - F(\mathbf{x}_i'\hat{\mathbf{a}}_N^{MV}).$$

- Si Monsieur *n* arrive, on lui pose des questions pour mesurer son $\mathbf{x}_n$, et on calcule $\hat{P}_N\left[y_i = 0 / \mathbf{x}_n\right] = 1 - F(\mathbf{x}_n'\hat{\mathbf{a}}_N^{MV})$. Monsieur *n* a peu de chance de se voir octroyer un crédit si « son » $1 - F(\mathbf{x}_n'\hat{\mathbf{a}}_N^{MV})$ est élevé.

Rmq. Le calcul de $1 - F(\mathbf{x}_i'\hat{\mathbf{a}}_N^{MV})$ est fondé sur la variable latente $y_i^* = \alpha_0 + \mathbf{b}_0'\mathbf{x}_i + u_i$. Les réponses de Monsieur *n* au questionnaire de « base » fournissent la valeur de $\mathbf{x}_n$.
La discussion avec Monsieur *n* et les détails sur la nature du projet de Monsieur *n* fournissent au banquier des éléments pour « prédire » u_n.

- On peut déterminer une règle de décision pour le banquier :

$$\delta(\mathbf{x}) : X \rightarrow \left| \begin{array}{l} 1 \text{ si l'individu "}\mathbf{x}\text{" est un bon client potentiel} \\ 0 \text{ si l'individu "}\mathbf{x}\text{" est un client à eviter} \end{array} \right.$$

- En utilisant le critère de l'erreur quadratique moyenne minimale :

$$\delta^*(\mathbf{x}) \equiv \arg\min_{\delta(.)} E\left[\left(y_i - \delta(\mathbf{x}_i) \right)^2 \middle/ \mathbf{x}_i = \mathbf{x} \right].$$

- Avec $E\left[y_i / \mathbf{x}_i \right] = P\left[y_i = 1 / \mathbf{x}_i \right] = F(\mathbf{x}_i' \mathbf{a}_0)$ on obtient :

$$\delta^*(\mathbf{x}) = \left| \begin{array}{l} 1 \text{ si } F_{u/\mathbf{x}}(\mathbf{x}'\mathbf{a}_0) > 1/2 \\ 0 \text{ si } F_{u/\mathbf{x}}(\mathbf{x}'\mathbf{a}_0) \leq 1/2 \end{array} \right. \quad \Rightarrow \quad \delta^*(\mathbf{x}) = \left| \begin{array}{l} 1 \text{ si } \mathbf{x}'\mathbf{a}_0 > 0 \\ 0 \text{ si } \mathbf{x}'\mathbf{a}_0 \leq 0 \end{array} \right.$$

- L'implication n'est valide que lorsque $f_{u/\mathbf{x}}(.)$ est symétrique par rapport (cas des lois normales centrées et des lois logistiques). Ce dernier résultat donne une justification à la construction des prédicteurs, les $\hat{y}_{i,N}$, utilisés ci-avant.

La méthode du référendum

Principe de la méthode

$$\left| \begin{array}{l} y_i = 1 \ \text{ si } i \text{ accepte } \delta_i \text{ pour le projet} \\ y_i = 0 \ \text{ si } i \text{ refuse } \delta_i \text{ pour le projet} \end{array} \right.$$

δ_i est la hausse d'impôts proposée à i et $\mathbf{x}_i$ décrit i

Formalisation économétrique

$$\left| \begin{array}{l} y_i = 1 \ \text{ si } \delta_i < c_i^* \\ y_i = 0 \ \text{ si } \delta_i \geq c_i^* \end{array} \right. \quad \text{et} \quad c_i^* = \mathbf{x}_i'\mathbf{a}_0 + u_i$$

c_i^* est le consentement maximal à payer de i pour le projet proposé

- L’intérêt de la méthode du référendum associée à la formalisation adoptée est qu’à partir d’un choix discret (simple d’un point de vue cognitif) on peut retirer des informations quantitatives quant aux préférences des individus pour le projet :

$\hat{c}^*_{i,N} \equiv \mathbf{x}'_i \hat{\mathbf{a}}^{MV}_N$ est un estimateur de la mesure « monétaire » de l’intérêt que i porte au projet

C’est une mesure du « ***bénéfice »*** (avant impôts !) ***de*** i ***si le projet était mené à terme***.

- A partir de $\hat{\mathbf{a}}^{MV}_N$ on peut :
  - Calculer un prédicteur $\hat{c}^*_{n,N} \equiv \mathbf{x}'_n \hat{\mathbf{a}}^{MV}_N$ du consentement à payer c^*_n d’un individu n de la population concernée mais non enquêté,
  - Calculer des bénéfices agrégés pour l’ensemble d’une population
  - Calculer une prédiction du résultat à un référendum ($\delta_i = \delta$) …

- On peut déterminer directement les log-vraisemblances correspondant au modèle :

$$\left| \begin{array}{l} y_i = 1 \text{ si } \delta_i < c_i^* \\ y_i = 0 \text{ si } \delta_i \geq c_i^* \end{array} \right. \quad \text{et} \quad c_i^* \,|\, (\mathbf{x}_i, \delta_i) \sim \mathcal{N}(\mathbf{x}_i'\mathbf{a}_0 ; \sigma_0^2)$$

  On obtient alors :

$$\boxed{L_N(\mathbf{a},\sigma) = \sum_{i=1}^{N} \left\{ y_i \times \ln\left[1 - \Phi\left[\frac{\delta_i - \mathbf{x}_i'\mathbf{a}}{\sigma}\right]\right] + (1 - y_i) \times \ln \Phi\left[\frac{\delta_i - \mathbf{x}_i'\mathbf{a}}{\sigma}\right] \right\}.}$$

- Les variations de δ_i sont ici exploitées pour estimer $\mathbf{a}_0$ et σ_0 séparément :
  - Les variations de $\mathbf{x}_i$ identifient $\mathbf{a}_0\sigma_0^{-1}$,
  - Les variations de δ_i identifient σ_0^{-1},
  - On a donc l'identification séparée de σ_0 et $\mathbf{a}_0$.

- Le modèle Probit estimé est équivalent au modèle Probit « standard » :

$$\begin{vmatrix} y_i = 1 & \text{si } y_i^* > 0 \\ y_i = 0 & \text{si } y_i^* \leq 0 \end{vmatrix} \quad \text{et} \quad y_i^* \mid \mathbf{z}_i \sim \mathcal{N}(\mathbf{z}_i'\boldsymbol{\alpha}_0 ; \sigma_0^2)$$

où :

$$y_i^* \equiv \mathbf{x}_i'\mathbf{a}_0 - \delta_i + u_i \equiv \mathbf{z}_i'\boldsymbol{\alpha}_0 + u_i = \sum_{k=1}^{K+1} \alpha_{k,0} z_{k,i} + u_i$$

avec

$$\alpha_{K+1,0} \equiv -1 \;\; \text{et} \;\; u_i \mid \mathbf{z}_i \sim \mathcal{N}(0, \sigma_0^2),$$

et donc avec la normalisation alternative à $\sigma_0 \equiv 1$.

- On peut calculer ici l'équivalent d'un R^2 :

$$R^2 = \frac{\hat{V}_N[\mathbf{x}_i'\hat{\mathbf{a}}_N - \delta_i]}{\hat{V}_N[\mathbf{x}_i'\hat{\mathbf{a}}_N - \delta_i] + \hat{\sigma}_N^2}$$

Train, K. and M. Weeks (2005). "Discrete choice models in preference space and WTP space" in *Applications of Simulation Methods in Environmental Resource Economics*, Alberini, A. and R. Scarpa (eds), Springer Publisher, Dordrecht, The Netherlands.

Train, K., R. Scarpa and M. Thiene (2008). "Utility in WTP Space: A Tool to Address Confounding Random Scale Effects in Destination Choice to the Alps" *Amer. J. Agr. Econ*. 90(4):994-1010.

6. Problèmes de specification

Remarque préliminaire. ***Effets d'intérêt***

Distinguer plusieurs cas d'utilisation de modèles de « variable dépendante dichotomique » :

(*i*) Prédiction de l'occurrence d'un évènement : ***problème de prédiction, i.e. d'ajustement***

(*ii*) Quantification des effets de déterminants observés sur un choix dichtomique : ***problème de prédiction, i.e. d'ajustement***

(*iii*) Utilisation des paramètres de la variable latente pour des calculs, tels que ceux des CAP de la méthode du référendum ou des calculs de surplus comme en économie industrielle appliquée : ***problème d'identification***.

Seul cas où le problème de la ***spécification du modèle de la variable latente*** est réellement important … économétrie « structurelle ».

- Dans de nombreuses études empiriques, le problème de l'économètre n'est que de « trouver » une spécification performante pour estimer :

  $$P[y_i = 1 | \mathbf{x}_i] \text{ : probabilité d'occurrence de l'évènement } y_i = 1$$

  i.e. de « trouver la bonne forme » de la fonction de répartition $F(.)$ dans :

  $$F(\mathbf{a}_0'\mathbf{x}_i) = P[y_i = 1 | \mathbf{x}_i],$$

  de « trouver le bon vecteur de variable explicatives » $\mathbf{x}_i$ et d'estimer $\mathbf{a}_0$.

- Puisqu'on a :

  $$E[y_i | \mathbf{x}_i] = P[y_i = 1 | \mathbf{x}_i],$$

  on est ici dans une logique de prédiction de y_i par $\mathbf{x}_i$, *i.e.* d'ajustement.

  - La structure latente de y_i n'a que peu d'intérêt ici.
  - On est ici dans une logique habituelle des statisticiens.

- ***Deux options*** dans ce cas :

  - Utilisation du *modèle de probabilité linaire*, *i.e*. de la projection linéaire de y_i sur $\mathbf{x}_i$, $EL[y_i \mid \mathbf{x}_i]$, ou d'une forme flexible pour $E[y_i \mid \mathbf{x}_i] = P[y_i = 1 \mid \mathbf{x}_i]$.
    - Estimation par les MC

  - On considère que $y_i \mid \mathbf{x}_i$ est une variable de Bernouilli de paramètre de propabilité dépendant de $\mathbf{x}_i$, *i.e.* :

$$y_i \sim \mathcal{B}_{p(\mathbf{x}_i)}.$$

  On choisit alors des modèles de $p(\mathbf{x}_i)$ de la forme :

$$p(\mathbf{x}_i) = F(\mathbf{a}_*'\mathbf{x}_i)$$

  où $F(.)$ est une fonction de répartition garantissant $p(\mathbf{x}_i) \in [0,1]$.
    - Estimation par le MV(C) ou le pseudo-MV(C)

- On cherche essentiellement à estimer :

$$\frac{\partial P[y_i = 1 | \mathbf{x}_i = \mathbf{x}]}{\partial x_{ki}} = a_{k,*} \times f(\mathbf{a}_*' \mathbf{x})$$

ou :

$$P[y_i = 1 | \mathbf{x}_i = \mathbf{x}^b] - P[y_i = 1 | \mathbf{x}_i = \mathbf{x}^a] = F(\mathbf{a}_*' \mathbf{x}^b) - F(\mathbf{a}_*' \mathbf{x}^a).$$

6.1. Les variables latentes en tant que variables d'intérêt

- Cependant, on s'intéresse également souvent à la variable latente $\tilde{y}_i$ ou aux variables latentes $(\tilde{U}_{1,i}, \tilde{U}_{0,i})$ sous-jacentes au PG de $y_i \mid \mathbf{x}_i$:

$$\begin{cases} y_i = 1 \text{ si } \tilde{y}_i > 0 \text{ ou } \tilde{U}_{1,i} > \tilde{U}_{0,i} \\ y_i = 0 \text{ si } \tilde{y}_i \leq 0 \text{ ou } \tilde{U}_{1,i} \leq \tilde{U}_{0,i} \end{cases}$$

- On s'intéressera au cas de la structuration par la variable latente $\tilde{y}_i$,
  - Le cas de la structuration par les utilités latentes sera développé dans les chapitres sur les modèles de choix de choix discret

- De fait, les modèles Probit ou Logit reposent sur des hypothèses assez restrictives quant au PG de $\tilde{y}_i \mid \mathbf{x}_i$:

$$\tilde{y}_i \mid \mathbf{x}_i \sim \mathcal{N}(\mathbf{x}_i'\mathbf{a}_0, \sigma_0^2) \quad \text{ou} \quad \tilde{y}_i \mid \mathbf{x}_i \sim \mathcal{L}(\mathbf{x}_i'\mathbf{a}_0, \gamma_0)$$

$$\boxed{\begin{array}{c}\tilde{y}_i = \mathbf{x}_i'\mathbf{a}_0 + u_i \\[2ex] u_i \mid \mathbf{x}_i \sim \mathcal{N}(0,\sigma_0^2) \quad \text{ou} \quad u_i \mid \mathbf{x}_i \sim \mathcal{L}(0,\gamma_0)\end{array}}$$

- La forme linéaire de $\tilde{y}_i = \mathbf{x}_i'\mathbf{a}_0 + u_i$ ne pose pas de problème outre mesure, cette forme est « flexible » pourvu qu'on choisisse « bien » $\mathbf{x}_i'\mathbf{a}_0$.

- Deux « gros » types de problèmes ici :
  - ***Endogénéité*** de certains éléments de $\mathbf{x}_i$, auquel cas :

$$E[\tilde{y}_i \mid \mathbf{x}_i] \neq \mathbf{x}_i'\mathbf{a}_0 \text{ car } E[u_i \mid \mathbf{x}_i] = \mu(\mathbf{x}_i) \neq 0.$$

  - ***Choix de*** $\mathcal{D}(u_i \mid \mathbf{x}_i)$, *e.g.* :

$$u_i \mid \mathbf{x}_i \sim \cancel{\mathcal{N}(0,\sigma_0^2)} \quad \Rightarrow \quad \ell_i(\sigma_0^{-1}\mathbf{a}_0) = \cancel{\Phi(\sigma_0^{-1} \times \mathbf{a}_0'\mathbf{x}_i)}\,.$$

- ***Hétéroscédasticité*** et ***famille paramétrique*** de $u_i \mid \mathbf{x}_i$

- ***Endogénéité*** de $\mathbf{x}_i$: chapitres dédiés dans la suite … problème spécifique

- ***Hypothèses sur*** $\mathcal{D}(u_i \mid \mathbf{x}_i)$, deux options :

  - Passage par des spécifications semi- ou non-paramétriques, pas ici.
    - ***Remarque***. On ne se passe pas complètement d'hypothèses sur $\mathcal{D}(u_i \mid \mathbf{x}_i)$, mais on en utilise de moins restrictives, *e.g*. symétrie de $\mathcal{D}(u_i \mid \mathbf{x}_i)$ par rapport à 0, …

  - ***Choix de familles de distribution alternatives*** à celles des lois normales ou logistiques : ***pas difficile*** (voir, *e.g*., Moon, 1988).
    - Distributions asymétriques en particulier, mais rare en pratique

  - Prise en compte de ***l'hétéroscédasticité*** potentielle ***de*** $u_i \mid \mathbf{x}_i$, *e.g*. :

    $u_i \mid \mathbf{x}_i \sim \mathcal{N}(0,\sigma_0^2) \quad \Rightarrow \quad u_i \perp\!\!\!\perp \mathbf{x}_i$ et, en particulier : $V[u_i \mid \mathbf{x}_i] = \sigma_0^2$.

    - Pas difficile mais quand-même rare en pratique

6.2. Le cas de l’hétéroscédasticité « exponentielle », Probit

- Cas du Probit ici
    - Logit et Logit Multinomial dans la suite
- En semi-paramétrique, *i.e.* pour une variable dépendante continue, la prise en compte de l’hétéroscédasticité n’engage que :
    - l’efficacité d’estimation

  et :
    - le calcul de la matrice de variance as. des estimateurs.
- L’hétéroscédasticité ne « gêne » pas la convergence des estimateurs des MC ou de la MM(G)

- *E.g.*, avec :

$$y_i = \mathbf{x}_i'\mathbf{a}_i + u_i \text{ avec } E[u_i \mid \mathbf{x}_i] = 0 \text{ et } V[u_i \mid \mathbf{x}_i] = h(\mathbf{x}_i^h;\boldsymbol{\delta}_0)^2 \neq \sigma_0^2.$$

  on a :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MCO} - \mathbf{a}_0) \xrightarrow[N\to+\infty]{L} \mathcal{N}\left(\mathbf{0}, E[\mathbf{x}_i\mathbf{x}_i']^{-1} E[h(\mathbf{x}_i^h;\boldsymbol{\delta}_0)^2 \mathbf{x}_i\mathbf{x}_i'] E[\mathbf{x}_i\mathbf{x}_i']^{-1}\right)$$

  et :

$$E[\mathbf{x}_i\mathbf{x}_i']^{-1} E[h(\mathbf{x}_i^h;\boldsymbol{\delta}_0)^2 \mathbf{x}_i\mathbf{x}_i'] E[\mathbf{x}_i\mathbf{x}_i']^{-1} \neq \sigma_0^2 E[\mathbf{x}_i\mathbf{x}_i']^{-1}.$$

- Pas la même chose, en général, en paramétrique avec variables latentes
  - Si les variables latentes d'un modèle paramétrique sont hétéroscédastiques alors cette hétéroscédasticité doit être (*i*) paramétrée et (*ii*) intégrée dans le modèle …
    - … sinon ***risque de non-convergence***
  - En utilisant le pseudo- ou quasi-MV (White, 1982, … ; Gouriéroux, Monfort et Trognon, 1984) on n'estime que …
    - … des ***pseudo-paramètres***.

Définition. ***Modèle Probit*** « ***hétéroscédastique*** »

$$\begin{cases} y_i = 1 \ \text{ si } \ \tilde{y}_i > 0 \\ y_i = 0 \ \text{ si } \ \tilde{y}_i \leq 0 \end{cases}$$

avec :

$$\tilde{y}_i \mid \mathbf{x}_i \sim \mathcal{N}\left(\alpha_0 + \tilde{\mathbf{x}}_i'\mathbf{b}_0, \sigma_0^2 h(\tilde{\mathbf{x}}_i^h;\boldsymbol{\delta}_0)^2\right) \text{ et } V[h(\tilde{\mathbf{x}}_i^h;\boldsymbol{\delta}_0)] \neq 0.$$

Contrainte de normalisation usuelle : $\sigma_0 \equiv 1$.

$\tilde{\mathbf{x}}_i^h$ est un sous-vecteur de $\tilde{\mathbf{x}}_i$ (ou $\tilde{\mathbf{x}}_i$ lui-même)

Exemple : $h(\tilde{\mathbf{x}}_i^h;\boldsymbol{\delta}_0) = \exp(\boldsymbol{\delta}_0'\tilde{\mathbf{x}}_i^h)$, ***et test de la variable ajoutée***

Modèle Probit à hétéroscédasticité exponentielle

$$\begin{cases} y_i = 1 \ \text{ si } \ \tilde{y}_i > 0 \\ y_i = 0 \ \text{ si } \ \tilde{y}_i \leq 0 \end{cases} \quad \text{avec} \quad \tilde{y}_i \mid \mathbf{x}_i \sim \mathcal{N}\left(\mathbf{x}_i'\mathbf{a}_0, \exp(2\times\boldsymbol{\delta}_0'\tilde{\mathbf{x}}_i^h)\right)$$

- Deux choses ici :
  - Discuter du caractère restrictif de $u_i \perp\!\!\!\perp \mathbf{x}_i$ …
  - … *via* un test d'hétéroscédasticité particulier, un exemple de ***test de spécification dit*** « ***de la variable ajoutée*** »
    - Voir Tauchen (1985) et Newey (1985)

- Dans ce cas on a :

$$h(\tilde{\mathbf{x}}_i^h;\boldsymbol{\delta}) = \exp(\boldsymbol{\delta}'\tilde{\mathbf{x}}_i^h) > 0,$$

ce qui est utile en pratique (et très souvent utilisé), et :

$$P[y_i = 1 | \mathbf{x}_i] = \Phi\left(\exp(\boldsymbol{\delta}_0'\tilde{\mathbf{x}}_i^h)^{-1} \times (\alpha_0 + \tilde{\mathbf{x}}_i'\mathbf{b}_0)\right)$$

et donc :

$$\boxed{\begin{aligned}\ln \ell_i(\mathbf{a},\boldsymbol{\delta}) = {} & y_i \times \ln\Phi\left(\exp(-\boldsymbol{\delta}'\tilde{\mathbf{x}}_i^h)\times(\mathbf{x}_i'\mathbf{a})\right) \\ & + (1-y_i)\times\ln\left[1-\Phi\left(\exp(-\boldsymbol{\delta}'\tilde{\mathbf{x}}_i^h)\times(\mathbf{x}_i'\mathbf{a})\right)\right].\end{aligned}}$$

- Bien entendu, avec :

MV*(*C*) *Probit « ***hétéroscédastique*** »

$$(\hat{\mathbf{a}}_N, \hat{\boldsymbol{\delta}}_N) \equiv \arg\max_{(\mathbf{a},\boldsymbol{\delta})} N^{-1} \sum_{i=1}^{N} \left\{ \begin{array}{c} y_i \times \ln \Phi\Big(\exp(-\boldsymbol{\delta}'\tilde{\mathbf{x}}_i^h) \times (\mathbf{x}_i'\mathbf{a})\Big) \\ +(1-y_i) \times \ln\Big[1 - \Phi\Big(\exp(-\boldsymbol{\delta}'\tilde{\mathbf{x}}_i^h) \times (\mathbf{x}_i'\mathbf{a})\Big)\Big] \end{array} \right\}$$

et :

MV*(*C*) *Probit standard (« ***homoscédastique*** »)

$$\tilde{\mathbf{a}}_N \equiv \arg\max_{\mathbf{a}} N^{-1} \sum_{i=1}^{N} \left\{ \begin{array}{c} y_i \times \ln \Phi(\mathbf{x}_i'\mathbf{a}) \\ +(1-y_i) \times \ln\big[1 - \Phi(\mathbf{x}_i'\mathbf{a})\big] \end{array} \right\} \neq \hat{\mathbf{a}}_N .$$

on a :

$$\tilde{\mathbf{a}}_N \neq \hat{\mathbf{a}}_N \quad \text{et} \quad p\lim_{N\to+\infty} \tilde{\mathbf{a}}_N \neq \mathbf{a}_0 = p\lim_{N\to+\infty} \hat{\mathbf{a}}_N$$

- Pour tester l'homoscédasticité du modèle il suffit de tester :

$$H_0 : \boldsymbol{\delta}_0 = \mathbf{0}$$

dans le modèle Probit augmenté (par $(\mathbf{x}_i'\mathbf{a}_0)\times\tilde{\mathbf{x}}_i^h$) :

$$\begin{cases} y_i = 1 \text{ si } \tilde{y}_i > 0 \\ y_i = 0 \text{ si } \tilde{y}_i \leq 0 \end{cases} \text{ avec } \begin{cases} \tilde{y}_i = \mathbf{x}_i'\mathbf{a}_0 - \boldsymbol{\delta}_0'\tilde{\mathbf{x}}_i^h(\mathbf{x}_i'\mathbf{a}_0) + e_i \\ e_i \mid \mathbf{x}_i \sim \mathcal{N}(0,1) \end{cases}$$

- On peut, dans $(\mathbf{x}_i'\mathbf{a}_0)\times\tilde{\mathbf{x}}_i^h$, remplacer $\mathbf{a}_0$ par un estimateur convergent $\tilde{\mathbf{a}}_N$, *e.g.*, celui du MV de $\mathbf{a}_0$ dans le modèle Probit standard et tester H_0 dans le modèle approché :

$$\begin{cases} y_i = 1 \text{ si } \tilde{y}_i > 0 \\ y_i = 0 \text{ si } \tilde{y}_i \leq 0 \end{cases} \text{ avec } \begin{cases} \tilde{y}_i = \mathbf{x}_i'\mathbf{a}_0 - \boldsymbol{\delta}_0'\tilde{\mathbf{x}}_i^h(\mathbf{x}_i'\tilde{\mathbf{a}}_N) + \tilde{e}_{i,N} \\ (p\lim_{N\to+\infty} \tilde{e}_{i,N}) \mid \mathbf{x}_i \sim \mathcal{N}(0,1) \end{cases}$$

sans correction pour la présence de l'estimateur $\tilde{\mathbf{a}}_N$: l'effet de son aléa disparaît sous H_0.

Procédure. ***Test de la variable ajoutée de l'hétéroscédasticité d'un Probit***

On veut tester :

$$H_0 : \boldsymbol{\delta}_0 = \mathbf{0},$$

i.e. le paramètre d'hétéroscédasticité du modèle Probit :

$$\begin{cases} y_i = 1 \text{ si } \tilde{y}_i > 0 \\ y_i = 0 \text{ si } \tilde{y}_i \le 0 \end{cases} \text{ avec } \tilde{y}_i \mid \mathbf{x}_i \sim \mathcal{N}\left(\mathbf{x}_i'\mathbf{a}_0, \exp(2\times\boldsymbol{\delta}_0'\tilde{\mathbf{x}}_i^h)\right)$$

Etape 1. Calcul de l'estimateur du MV contraint de $(\mathbf{a}_0, \boldsymbol{\delta}_0)$, $(\mathbf{a}_0, \mathbf{0})$:

$$\tilde{\mathbf{a}}_N \equiv \arg\max_{\mathbf{a}} N^{-1}\sum_{i=1}^{N}\begin{Bmatrix} y_i \times \ln\Phi(\mathbf{x}_i'\mathbf{a}) \\ +(1-y_i)\times\ln\left[1-\Phi(\mathbf{x}_i'\mathbf{a})\right] \end{Bmatrix},$$

i.e. l'estimateur du MV contraint de $(\mathbf{a}_0, \boldsymbol{\delta}_0)$ dans le Probit « standard » :

$$\begin{cases} y_i = 1 \text{ si } \tilde{y}_i > 0 \\ y_i = 0 \text{ si } \tilde{y}_i \le 0 \end{cases} \text{ avec } \tilde{y}_i \mid \mathbf{x}_i \sim \mathcal{N}(\mathbf{x}_i'\mathbf{a}_0, 1).$$

Etape 2. Calcul de l'estimateur du MV de $(\mathbf{a}_0, \boldsymbol{\delta}_0)$:

$$(\breve{\mathbf{a}}_N, \breve{\boldsymbol{\delta}}_N) \equiv \arg\max_{(\mathbf{a},\boldsymbol{\delta})} N^{-1} \sum_{i=1}^{N} \left\{ \begin{array}{c} y_i \times \ln \Phi\left(\mathbf{x}_i'\mathbf{a} - (\mathbf{x}_i'\tilde{\mathbf{a}}_N)\boldsymbol{\delta}'\tilde{\mathbf{x}}_i^h \right) \\ +(1-y_i) \times \ln\left[1 - \Phi\left(\mathbf{x}_i'\mathbf{a} - (\mathbf{x}_i'\tilde{\mathbf{a}}_N)\boldsymbol{\delta}'\tilde{\mathbf{x}}_i^h \right) \right] \end{array} \right\},$$

i.e. l'estimateur du MV de $(\mathbf{a}_0, \boldsymbol{\delta}_0)$ dans le Probit standard « augmenté » de la variable $(\mathbf{x}_i'\mathbf{a}_0) \times \tilde{\mathbf{x}}_i^h$:

$$\begin{cases} y_i = 1 \text{ si } \tilde{y}_i > 0 \\ y_i = 0 \text{ si } \tilde{y}_i \leq 0 \end{cases} \text{ avec } \tilde{y}_i \mid \mathbf{x}_i \sim \mathcal{N}\left(\mathbf{x}_i'\mathbf{a}_0 - \boldsymbol{\delta}_0'\tilde{\mathbf{x}}_i^h (\mathbf{x}_i'\mathbf{a}_0), 1 \right).$$

Etape 3. Procédure standard de test de :

$$H_0 : \boldsymbol{\delta}_0 = \mathbf{0},$$

i.e. en utilisant le test de Wald usuel.

(**3.a**) Si H_0 acceptée, utiliser $\tilde{\mathbf{a}}_N$. Il est as. efficace pour $\mathbf{a}_0$.

(**3.b**) Si H_0 rejetée, calculer :

$$(\hat{\mathbf{a}}_N, \hat{\boldsymbol{\delta}}_N) \equiv \arg\max_{(\mathbf{a},\boldsymbol{\delta})} N^{-1} \sum_{i=1}^{N} \left\{ \begin{array}{c} y_i \times \ln \Phi\left(\exp(-\boldsymbol{\delta}' \tilde{\mathbf{x}}_i^h) \times (\mathbf{x}_i' \mathbf{a})\right) \\ +(1-y_i) \times \ln\left[1 - \Phi\left(\exp(-\boldsymbol{\delta}' \tilde{\mathbf{x}}_i^h) \times (\mathbf{x}_i' \mathbf{a})\right)\right] \end{array} \right\}.$$

Le test de Wald de $H_0 : \boldsymbol{\delta}_0 = \mathbf{0}$ dans le Probit « augmenté » est as. équivalent au test du Score (ou du Multiplicateur de Lagrange) usuel de $\boldsymbol{\delta}_0 = \mathbf{0}$.

- Cette procédure de test a pour origine les liens entre la forme de la fonction de score du Probit standard et celle du Probit « à hétéroscédasticité exponentielle ».
    - Pas vraiment important ici, mais quand-même voir ci-après.

- En fait, la principale implication à retirer de l'existence de ce test est que, *dixit* J. Woodridge (2010) :

  « *L'approche par les variables ajoutées révèle que le test du score de l'hétéroscédasticité du terme d'erreur de la variable latente ne peut être distingué d'un test d'une forme particulière d'effets croisés des variables explicatives.*

  On aurait pu arriver directement à ce constat, sans modèle de $V[u_i \mid \mathbf{x}_i]$. »

- L'analyse de ce constat s'appuie sur deux remarques :
  - La première est liée au calcul des estimateurs du MV(C),
  - La seconde est liée à la forme de $E[y_i \mid \mathbf{x}_i] = P[y_i = 1 \mid \mathbf{x}_i]$

Choix dichotomique, MV(C), MM et MC : spécification correcte

- On suppose ici que la vraisemblance $\ell_i(\mathbf{a}) = F(\mathbf{x}_i'\mathbf{a})^{y_i} \times \left(1 - F(\mathbf{x}_i'\mathbf{a})\right)^{1-y_i}$ est correctement spécifiées, *i.e*. que :

$$\boxed{P[y_i = 1 \mid \mathbf{x}_i] = F(\mathbf{x}_i'\mathbf{a}_0) \text{ et que } \mathbf{a}_0 \equiv \arg\max_{\mathbf{a}} E\left[\ln \ell_i(a)\right]}$$

- Les CO1 du programme définissant $\hat{\mathbf{a}}_N^{MV}$:

$$\boxed{\hat{\mathbf{a}}_N^{MV} \equiv \arg\max_{\mathbf{a}} N^{-1} \sum_{i=1}^{N} \left\{ y_i \times \ln F(\mathbf{x}_i'\mathbf{a}) + (1 - y_i) \times \ln\left[1 - F(\mathbf{x}_i'\mathbf{a})\right] \right\}}$$

sont données par :

$$N^{-1} \sum_{i=1}^{N} \mathbf{x}_i \times \frac{f(\mathbf{x}_i'\hat{\mathbf{a}}_N^{MV})}{\left(1 - F(\mathbf{x}_i'\hat{\mathbf{a}}_N^{MV})\right) \times F(\mathbf{x}_i'\hat{\mathbf{a}}_N^{MV})} \times \left(y_i - F(\mathbf{x}_i'\hat{\mathbf{a}}_N^{MV})\right) = \mathbf{0}.$$

- Ces CO1 correspondent également à celle de l'estimateur des MCG du modèle de régression à termes d'erreur hétéroscédastiques déduit de l'espérance conditionnelle :

$$\boxed{E[y_i \mid \mathbf{x}_i] = F(\mathbf{x}_i'\mathbf{a}_0) = P[y_i = 1 \mid \mathbf{x}_i]}$$

- En effet, on a :

$$y_i = F(\mathbf{x}_i'\mathbf{a}_0) + v_i \quad \text{avec} \quad E[v_i \mid \mathbf{x}_i] = 0 \quad \text{et} \quad \begin{cases} v_i = 1 - F(\mathbf{x}_i'\mathbf{a}_0) & \text{si } y_i = 1 \\ v_i = -F(\mathbf{x}_i'\mathbf{a}_0) & \text{si } y_i = 0 \end{cases}$$

- On a alors :

$$V[v_i \mid \mathbf{x}_i] = F(\mathbf{x}_i'\mathbf{a}_0) \times \left(1 - F(\mathbf{x}_i'\mathbf{a}_0)\right) \equiv w(\mathbf{x}_i'\mathbf{a}_0)^2.$$

- Les CO1 du programme définissant $\hat{\mathbf{a}}_N^{MCG}$, *i.e.* :

$$\boxed{\hat{\mathbf{a}}_N^{MCG} \equiv \arg\min_{\mathbf{a}} N^{-1} \sum_{i=1}^{N} w(\mathbf{x}_i'\mathbf{a})^{-2} \times \left(y_i - F(\mathbf{x}_i'\mathbf{a}) \right)^2}$$

caractérisent à la fois $\hat{\mathbf{a}}_N^{MCG}$ et $\hat{\mathbf{a}}_N^{MV}$.

- La spécification correcte de $E[y_i \mid \mathbf{x}_i] = F(\mathbf{x}_i'\mathbf{a}_0) = P[y_i = 1 \mid \mathbf{x}_i]$ garantit également que :

$$\mathbf{a}_0 \equiv \arg\min_{\mathbf{a}} E\left[w(\mathbf{x}_i)^{-1} \left(y_i - F(\mathbf{x}_i'\mathbf{a}) \right)^2 \right]$$

  pour tout fonction de « poids » $w(\mathbf{x}_i) > 0$.

Définition. ***Résidu standardisé***

Le terme :

$$\frac{y_i - F(\mathbf{x}_i'\mathbf{a})}{w(\mathbf{x}_i'\mathbf{a})} = \frac{y_i - F(\mathbf{x}_i'\mathbf{a})}{\sqrt{F(\mathbf{x}_i'\mathbf{a}_0) \times \left(1 - F(\mathbf{x}_i'\mathbf{a}_0)\right)}}$$

est nommé résidu standardisé du modèle $y_i = F(\mathbf{x}_i'\mathbf{a}_0) + v_i$ avec $E[v_i \mid \mathbf{x}_i] = 0$.

- Les estimateurs $\hat{\mathbf{a}}_N^{MCG} = \hat{\mathbf{a}}_N^{MV}$ peuvent être interprétés comme des ***estimateurs de la MM*** fondés sur la ***condition d'orthogonalité « Chamberlain optimale »*** :

$$E\left[\mathbf{x}_i f(\mathbf{x}_i'\mathbf{a}_0) w(\mathbf{x}_i'\mathbf{a}_0)^{-2}\left(y_i - F(\mathbf{x}_i'\mathbf{a}_0)\right)\right] = \mathbf{0},$$

  dérivée de la ***condition de moment conditionnelle*** :

$$E\left[y_i - F(\mathbf{x}_i'\mathbf{a}_0) \mid \mathbf{x}_i\right] = 0$$

  puisque :

$$w(\mathbf{x}_i'\mathbf{a}_0)^2 \equiv V[v_i \mid \mathbf{x}_i] \text{ et } E\left[\partial F(\mathbf{x}_i'\mathbf{a}_0)/\partial \mathbf{a} \mid \mathbf{x}_i\right] = \mathbf{x}_i f(\mathbf{x}_i'\mathbf{a}_0).$$

- On a alors :

$$\boxed{\hat{\mathbf{a}}_N^{MM} \text{ solution en } \mathbf{a} \text{ de } N^{-1}\sum_{i=1}^{N} \mathbf{x}_i f(\mathbf{x}_i'\mathbf{a}) w(\mathbf{x}_i'\mathbf{a})^{-2}\left(y_i - F(\mathbf{x}_i'\mathbf{a})\right) = \mathbf{0}}$$

Remarque. $\hat{\mathbf{a}}_N^{MM}$ n'est unique que si $L_N(\mathbf{a}) \equiv N^{-1}\sum_{i=1}^{N} \ln \ell_i(\mathbf{a})$ est strictement concave en $\mathbf{a}$.

Choix dichotomique, MV(C), MM et MC, et problèmes de spécification

Spécification de $E[y_i \mid \mathbf{x}_i] = P[y_i = 1 \mid \mathbf{x}_i]$ ***et spécification de*** $\mathcal{D}(u_i \mid \mathbf{x}_i)$

Structure latente d'un modèle de choix dichotomique

$$\begin{cases} y_i = 1 \text{ si } \tilde{y}_i > 0 \\ y_i = 0 \text{ si } \tilde{y}_i \leq 0 \end{cases} \text{ avec } \tilde{y}_i = \mathbf{x}_i'\mathbf{a}_0 + u_i \text{ et } E[u_i \mid \mathbf{x}_i] = E[u_i] \equiv 0$$

Plusieurs cas possibles par rapport à la spécification, ou de mauvaise, spécification de $\mathcal{D}(u_i \mid \mathbf{x}_i)$

On suppose ici que la « véritable » $\mathcal{D}(u_i \mid \mathbf{x}_i)$ est caractérisée par la fonction de répartition $G_0(u; \mathbf{x}_i)$

Modèle de choix dichotomique paramétrique standard,
i.e. à variable latente dont le terme d'erreur est homoscédastique

Hyp 1 : $G_0(u;\mathbf{x}_i) = F(u)$ où $F(.)$ est de forme connue, *i.e.* $G_0(u;\mathbf{x}_i)$, et donc $\mathcal{D}(u_i \,|\, \mathbf{x}_i)$, sont connue.

Hyp 2 : On a $\mathcal{D}(u_i \,|\, \mathbf{x}_i) = \mathcal{D}(u_i)$, *i.e.* u_i et $\mathbf{x}_i$ sont indépendantes.

Dans ce cas ***les résultats relatifs à l'inférence par le MV***(***C***) s'appliquent, et l'estimateur du MV(C) de $\mathbf{a}_0$ est as. efficace.

Utilisé dans 95% des cas

Remarque

On ne s'intéresse qu'à un type de problème de spécification ici, celui du choix de la famille paramétrique de $G_0(u;\mathbf{x}) = P[y_i = 1 | \mathbf{x}_i = \mathbf{x}]$.

On ne considère pas le problème de l'endogénéité d'un ou de plusieurs éléments de $\mathbf{x}_i$ dans le modèle de y_i.

En effet, dans ce cas le problème est mal posé « dès le départ » car si on discute de problème d'endogénéité c'est que $\mathcal{D}(y_i | \mathbf{x}_i)$ n'est pas la distribution d'intérêt.

Ici on s'intéresse au cas où on cherche à estimer $G_0(u;\mathbf{x}) = P[y_i = 1 | \mathbf{x}_i = \mathbf{x}]$, ou des éléments dérivés de cette fonction de probabilité, sachant qu'on a du mal à déterminer la loi de $\mathcal{D}(y_i | \mathbf{x}_i)$.

On se concentre ici sur le problème de l'hétéroscédasticité de $\tilde{y}_i | \mathbf{x}_i$.

Modèle de choix dichotomique paramétrique à variable latente dont le terme d'erreur est hétéroscédastique

Hyp 1 : $G_0(u;\mathbf{x}_i) = F\left(h(\mathbf{x}_i^h;\boldsymbol{\delta}_0)^{-1} \times u\right)$ où $F(.)$ et $h(.)$ sont de formes connues,

i.e. $G_0(u;\mathbf{x}_i)$, et donc $\mathcal{D}(u_i \mid \mathbf{x}_i)$, sont connues.

Hyp 2 : On a $V[u_i \mid \mathbf{x}_i] \propto h(\mathbf{x}_i^h;\boldsymbol{\delta}_0)^2$, *i.e.* u_i est hétéroscédastique$\mid \mathbf{x}_i$.

Dans ce cas ***les résultats relatifs à l'inférence par le MV***(***C***) s'appliquent, et l'estimateur du MV(C) de $\mathbf{a}_0$ est as. efficace.

L'estimateur du MV(C) de $\mathbf{a}_0$ utilisant $F(u)$, i.e. négligeant l'hétéroscédasticité des u_i, n'est pas convergent.

Avec :

$$\tilde{\mathbf{a}}_N \equiv \arg\max_{\mathbf{a}} N^{-1}\sum_{i=1}^{N} \ln \ell_i^m(\mathbf{a})$$

où :

$$\ln \ell_i^m(\mathbf{a}) \equiv y_i \times \ln F(\mathbf{a}'\mathbf{x}_i) + (1-y_i)\times \ln\left(1-F(\mathbf{a}'\mathbf{x}_i)\right)$$

on a :

$$p\lim_{N\to+\infty} \tilde{\mathbf{a}}_N = \mathbf{a}_* \equiv \arg\max_{\mathbf{a}} E\left[\ln \ell_i^m(\mathbf{a})\right] \neq \mathbf{a}_0 = \arg\max_{\mathbf{a}} E\left[\ln \ell_i(\mathbf{a},\boldsymbol{\delta}_0)\right].$$

On a également :

$$\mathbf{a}_* = \arg\min_{\mathbf{a}} E\left[w(\mathbf{x}_i'\mathbf{a})^{-2} \times \left(y_i - F(\mathbf{x}_i'\mathbf{a})\right)^2 \right].$$

et :

$$\mathbf{a}_* \text{ solution en } \mathbf{a} \text{ de } E\left[\mathbf{x}_i f(\mathbf{x}_i'\mathbf{a}) w(\mathbf{x}_i'\mathbf{a})^{-2}\left(y_i - F(\mathbf{x}_i'\mathbf{a})\right)\right] = \mathbf{0}$$

mais ça a peu d'intérêt ici.

Rôle des « variables ajoutées »

Modèle de choix dichotomique à variable latente hétéroscédastique

$$\begin{cases} y_i = 1 \text{ si } \tilde{y}_i > 0 \\ y_i = 0 \text{ si } \tilde{y}_i \leq 0 \end{cases} \text{ avec } \tilde{y}_i = \mathbf{x}_i'\mathbf{a}_0 + u_i \text{ et } E[u_i \mid \mathbf{x}_i] = E[u_i] \equiv 0$$

La distribution de $u_i \mid \mathbf{x}_i$ est caractérisée par la fonction de répartition conditionnelle : $G_0(u;\mathbf{x}_i) = F\left(h(\mathbf{x}_i^h;\boldsymbol{\delta}_0)^{-1} \times u\right)$,

avec $V[h(\mathbf{x}_i^h;\boldsymbol{\delta}_0)] \neq 0$ et $h(\mathbf{x}_i^h;\boldsymbol{\delta}) > 0$

« Modèle » de variable latente « augmenté »

$$\tilde{y}_i = \mathbf{a}_+'\mathbf{x}_i + \boldsymbol{\gamma}'(\mathbf{x}_i \otimes \mathbf{x}_i^h) + e_i \text{ et } E[e_i] \equiv 0$$

Pas vraiment un modèle, juste un outil car $\mathbf{a} = \mathbf{a}_0$ et $\boldsymbol{\gamma} = \mathbf{0}$ ici.

Log-vraisemblance correcte

$$\ln \ell_i(\mathbf{a},\boldsymbol{\delta}) \equiv y_i \times \ln F\left(h(\mathbf{x}_i^h;\boldsymbol{\delta})^{-1} \times \mathbf{a}'\mathbf{x}_i\right) + (1-y_i)\times \ln\left[1 - F\left(h(\mathbf{x}_i^h;\boldsymbol{\delta})^{-1} \times \mathbf{a}'\mathbf{x}_i\right)\right]$$

Log-vraisemblance « ***homoscédastique*** »

$$\ln \ell_i^m(\mathbf{a}) \equiv y_i \times \ln F(\mathbf{a}'\mathbf{x}_i) + (1-y_i)\times \ln\left(1 - F(\mathbf{a}'\mathbf{x}_i)\right)$$

Log-vraisemblance « augmentée »

$$\ln \ell_i^+(\mathbf{a},\boldsymbol{\beta}) \equiv y_i \times \ln F\left(\mathbf{a}'\mathbf{x}_i + \boldsymbol{\gamma}'(\mathbf{x}_i \otimes \mathbf{x}_i^h)\right) + (1-y_i)\times \ln\left[1 - F\left(\mathbf{a}'\mathbf{x}_i + \boldsymbol{\gamma}'(\mathbf{x}_i \otimes \mathbf{x}_i^h)\right)\right]$$

$\ln \ell_i^m(\mathbf{a})$ est un cas particulier de $\ln \ell_i^+(\mathbf{a},\boldsymbol{\gamma})$, avec $\boldsymbol{\gamma} = \mathbf{0}$

Probit si $F(.) = \Phi(.)$

- On notera, en supposant l'unicité des programmes d'optimisation :

$$(\mathbf{a}_0,\boldsymbol{\delta}_0) = \arg\max_{(\mathbf{a},\boldsymbol{\delta})} E[\ln \ell_i(\mathbf{a},\boldsymbol{\delta})]$$

$$\mathbf{a}_* \equiv \arg\max_{\mathbf{a}} E[\ln \ell_i^m(\mathbf{a})]$$

$$(\mathbf{a}_+,\boldsymbol{\gamma}_+) \equiv \arg\max_{(\mathbf{a},\boldsymbol{\gamma})} E[\ln \ell_i^+(\mathbf{a},\boldsymbol{\gamma})]$$

- Sous certaines conditions de régularité on a :

$$p\lim_{N\to+\infty}(\hat{\mathbf{a}}_N,\hat{\boldsymbol{\delta}}_N) = (\mathbf{a}_0,\boldsymbol{\delta}_0) \quad \text{où} \quad (\hat{\mathbf{a}}_N,\hat{\boldsymbol{\delta}}_N) \equiv \arg\max_{(\mathbf{a},\boldsymbol{\delta})} N^{-1}\sum_{i=1}^{N} \ln \ell_i(\mathbf{a},\boldsymbol{\delta})$$

$$p\lim_{N\to+\infty}\tilde{\mathbf{a}}_N = \mathbf{a}_0 \quad \text{où} \quad \tilde{\mathbf{a}}_N \equiv \arg\max_{\mathbf{a}} N^{-1}\sum_{i=1}^{N} \ln \ell_i^m(\mathbf{a})$$

$$p\lim_{N\to+\infty}(\breve{\mathbf{a}}_N,\breve{\boldsymbol{\gamma}}_N) = (\mathbf{a}_+,\boldsymbol{\gamma}_+) \quad \text{où} \quad (\breve{\mathbf{a}}_N,\breve{\boldsymbol{\gamma}}_N) \equiv \arg\max_{(\mathbf{a},\boldsymbol{\gamma})} N^{-1}\sum_{i=1}^{N} \ln \ell_i^+(\mathbf{a},\boldsymbol{\gamma})$$

- Comparer les spécifications de $E[y_i \mid \mathbf{x}_i] = P[y_i = 1 \mid \mathbf{x}_i]$ donne :

Sous les hypothèses posées on a :
(i) $\mathbf{a}_* \neq \mathbf{a}_0$
(ii) $\mathbf{a}_+ \neq \mathbf{a}_0$ et, probablement $\mathbf{a}_+ \neq \mathbf{a}_*$ et $\boldsymbol{\gamma}_+ \neq \mathbf{0}$.

- L'inégalité de Kullback-Leibler et les hypothèses posées impliquent que :

$$\boxed{E[\ln \ell_i(\mathbf{a}_0, \boldsymbol{\delta}_0)] > E[\ln \ell_i^m(\mathbf{a}_*)] \quad \text{et} \quad E[\ln \ell_i(\mathbf{a}_0, \boldsymbol{\delta}_0)] \geq E[\ln \ell_i^+(\mathbf{a}_+, \boldsymbol{\gamma}_+)].}$$

- Le fait que $\ln \ell_i^m(\mathbf{a}) = \ln \ell_i^+(\mathbf{a}, \mathbf{0})$ donne :

$$\boxed{E[\ln \ell_i(\mathbf{a}_0, \boldsymbol{\delta}_0)] \geq E[\ln \ell_i^+(\mathbf{a}_+, \boldsymbol{\gamma}_+)] \geq E[\ln \ell_i^m(\mathbf{a}_*)].}$$

Interprétation en termes de prédiction de $P[y_i = 1 | \mathbf{x}_i]$

$$P[y_i = 1 | \mathbf{x}_i] = F\left(h(\mathbf{x}_i^h; \boldsymbol{\delta}_0)^{-1} \times \mathbf{a}_0' \mathbf{x}_i\right)$$

Le terme $F(\mathbf{a}_*' \mathbf{x}_i)$ est le meilleur prédicteur de $P[y_i = 1 | \mathbf{x}_i]$ par une fonction de la forme $F(\mathbf{a}'\mathbf{x})$ au sens du maximum de l'information de Kullback-Leibler (logique du pseudo-MV)

Le terme $F\left(\mathbf{a}_+' \mathbf{x}_i + \boldsymbol{\gamma}_+' (\mathbf{x}_i \otimes \mathbf{x}_i^h)\right)$ est le meilleur prédicteur de $P[y_i = 1 | \mathbf{x}_i]$ par une fonction de la forme $F\left(\mathbf{a}' \mathbf{x}_i + \boldsymbol{\gamma}' (\mathbf{x}_i \otimes \mathbf{x}_i^h)\right)$ au sens du maximum de l'information de Kullback-Leibler (logique du pseudo-MV)

$F\left(\mathbf{a}_+' \mathbf{x}_i + \boldsymbol{\gamma}_+' (\mathbf{x}_i \otimes \mathbf{x}_i^h)\right)$ est un meilleur prédicteur de $P[y_i = 1 | \mathbf{x}_i]$ que $F(\mathbf{a}_*' \mathbf{x}_i)$. Il utilise une « fonction support » plus générale.

Le terme $F\left(\mathbf{a}'_{+}\mathbf{x}_i+\boldsymbol{\gamma}'_{+}(\mathbf{x}_i\otimes\mathbf{x}_i^h)\right)$ est le meilleur prédicteur de $F\left(h(\mathbf{x}_i^h;\boldsymbol{\delta}_0)^{-1}\times\mathbf{a}'_0\mathbf{x}_i\right)$ par une fonction de la forme $F\left(\mathbf{a}'\mathbf{x}_i+\boldsymbol{\gamma}'(\mathbf{x}_i\otimes\mathbf{x}_i^h)\right)$ au sens du maximum de l'information de Kullback-Leibler.

I.e., $\mathbf{a}'_{+}\mathbf{x}_i+\boldsymbol{\gamma}'_{+}(\mathbf{x}_i\otimes\mathbf{x}_i^h)$ est définie comme une approximation de $h(\mathbf{x}_i^h;\boldsymbol{\delta}_0)^{-1}\times\mathbf{a}'_0\mathbf{x}_i$ au sens du critère d'ajustement spécifique du MV(C).

Problème d'interprétation si « ***doute général sur la spécification*** »

Si on doute également sur les formes de :

$\mathbf{a}'_0\mathbf{x}_i$, $h(\mathbf{x}_i^h;\boldsymbol{\delta}_0)$ et $F(.)$,

alors on ne sait pas trop ce qu'on teste avec le test de la variable ajoutée.

Ici on ne s'est posé que la question suivante : a-t-on $h(\mathbf{x}_i^h;\boldsymbol{\delta}_0)=cst$ ou non ?

Annexe. ***Eléments d'analyse du test de la variable ajoutée dans le Probit « à hétéroscédasticité exponentielle »***

Définition. *Modèle Probit* « ***à hétéroscédasticité exponentielle*** »

$$\begin{cases} y_i = 1 \text{ si } \tilde{y}_i > 0 \\ y_i = 0 \text{ si } \tilde{y}_i \leq 0 \end{cases}$$

avec :

$$\tilde{y}_i \,|\, \mathbf{x}_i \sim \mathcal{N}\left(\alpha_0 + \tilde{\mathbf{x}}_i'\mathbf{b}_0, \sigma_0^2 \exp(\boldsymbol{\delta}_0'\tilde{\mathbf{x}}_i^h)^2\right) \text{ et } V[h(\tilde{\mathbf{x}}_i^h; \boldsymbol{\delta}_0)] \neq 0.$$

Contrainte de normalisation usuelle : $\sigma_0 \equiv 1.$

$\tilde{\mathbf{x}}_i^h$ est un sous-vecteur de $\tilde{\mathbf{x}}_i$ (ou $\tilde{\mathbf{x}}_i$ lui-même)

$$\begin{aligned} \ln \ell_i(\mathbf{a}, \boldsymbol{\delta}) = {} & y_i \times \ln \Phi\left(\exp(-\boldsymbol{\delta}'\tilde{\mathbf{x}}_i^h) \times (\mathbf{x}_i'\mathbf{a})\right) \\ & + (1 - y_i) \times \ln\left[1 - \Phi\left(\exp(-\boldsymbol{\delta}'\tilde{\mathbf{x}}_i^h) \times (\mathbf{x}_i'\mathbf{a})\right)\right] \end{aligned}$$

- En notant :

$$h_i(-\boldsymbol{\delta}) \equiv \exp(-\boldsymbol{\delta}\tilde{\mathbf{x}}_i^h),$$

$$P_i^h(\mathbf{x}_i'\mathbf{a},\boldsymbol{\delta}) \equiv \Phi\left(h_i(-\boldsymbol{\delta})\times(\mathbf{x}_i'\mathbf{a})\right)$$

$$g_i^h(\mathbf{x}_i'\mathbf{a},\boldsymbol{\delta}) \equiv \varphi\left(h_i(-\boldsymbol{\delta})\times(\mathbf{x}_i'\mathbf{a})\right)\times\frac{y_i - P_i^h(\mathbf{x}_i'\mathbf{a},\boldsymbol{\delta})}{\left(1-P_i^h(\mathbf{x}_i'\mathbf{a},\boldsymbol{\delta})\right)\times P_i^h(\mathbf{x}_i'\mathbf{a},\boldsymbol{\delta})}$$

on montre, après quelques calculs simples, que les CO1 du programme définissant $(\mathbf{a}_0,\boldsymbol{\delta}_0)$:

$$(\mathbf{a}_0,\boldsymbol{\delta}_0) \equiv \arg\max_{(\mathbf{a},\boldsymbol{\delta})} E\left[y_i \times \ln P_i^h(\mathbf{x}_i'\mathbf{a},\boldsymbol{\delta}) + (1-y_i)\times\ln\left(1-P_i^h(\mathbf{x}_i'\mathbf{a},\boldsymbol{\delta})\right)\right]$$

donnent les équations de score du modèle Probit « hétéroscédastique considéré ici :

$$\boxed{\begin{cases} \text{Score en } \mathbf{a} : E\left[\mathbf{x}_i \times h_i(-\boldsymbol{\delta}_0) \times g_i^h(\mathbf{x}_i'\mathbf{a}_0, \boldsymbol{\delta}_0) \right] = \mathbf{0} \\ \text{Score en } \boldsymbol{\delta} : E\left[-\tilde{\mathbf{x}}_i^h \times h_i(-\boldsymbol{\delta}_0) \times (\mathbf{x}_i'\mathbf{a}_0) \times g_i^h(\mathbf{x}_i'\mathbf{a}_0, \boldsymbol{\delta}_0) \right] = \mathbf{0} \end{cases}}$$

car :

$$\text{Score en } \mathbf{a} : \frac{\partial \ln \ell_i(\mathbf{a}, \boldsymbol{\delta})}{\partial \mathbf{a}} = \mathbf{x}_i \times h_i(-\boldsymbol{\delta}) \times g_i^h(\mathbf{x}_i'\mathbf{a}, \boldsymbol{\delta})$$

et :

$$\text{Score en } \boldsymbol{\delta} : \frac{\partial \ln \ell_i(\mathbf{a}, \boldsymbol{\delta})}{\partial \boldsymbol{\delta}} = -\tilde{\mathbf{x}}_i^h \times h_i(-\boldsymbol{\delta}) \times (\mathbf{x}_i'\mathbf{a}) \times g_i^h(\mathbf{x}_i'\mathbf{a}, \boldsymbol{\delta}).$$

- Les CO1 (et leurs $p\lim_{N \to +\infty}$) du programme de maximisation :

$$(\breve{\mathbf{a}}_N, \breve{\boldsymbol{\delta}}_N) \equiv \arg\max_{(\mathbf{a},\boldsymbol{\delta})} N^{-1} \sum_{i=1}^{N} \left\{ \begin{array}{c} y_i \times \ln \Phi\left(\mathbf{x}_i'\mathbf{a} - (\mathbf{x}_i'\tilde{\mathbf{a}}_N)\boldsymbol{\delta}'\tilde{\mathbf{x}}_i^h \right) \\ +(1-y_i) \times \ln\left[1 - \Phi\left(\mathbf{x}_i'\mathbf{a} - (\mathbf{x}_i'\tilde{\mathbf{a}}_N)\boldsymbol{\delta}'\tilde{\mathbf{x}}_i^h \right) \right] \end{array} \right\}$$

utilisent des fonctions de score identiques dès lors que $\boldsymbol{\delta}_0 = \mathbf{0}$.

- Pour cela il suffit de remarquer que (*i*) :

$$(\breve{\mathbf{a}}_N,\breve{\boldsymbol{\delta}}_N)\equiv\arg\max_{(\mathbf{a},\boldsymbol{\delta})} N^{-1}\sum_{i=1}^{N}\left\{\begin{array}{c} y_i\times\ln P_i^h\left(\mathbf{x}_i'\mathbf{a}-(\mathbf{x}_i'\tilde{\mathbf{a}}_N)\boldsymbol{\delta}'\tilde{\mathbf{x}}_i^h,\mathbf{0}\right)\\ +(1-y_i)\times\ln\left[1-P_i^h\left(\mathbf{x}_i'\mathbf{a}-(\mathbf{x}_i'\tilde{\mathbf{a}}_N)\boldsymbol{\delta}'\tilde{\mathbf{x}}_i^h,\mathbf{0}\right)\right]\end{array}\right\}$$

et que (*i*) on retrouve le modèle Probit standard à partir du Probit « hétéroscédastique » si $\boldsymbol{\delta}_0=\mathbf{0}$ car (*ii*) $h_i(\mathbf{0})=1$.

- De fait, les CO1 du programme de maximisation défini ci-dessus sont données par :

$$\boxed{\left\{\begin{array}{l} \text{CO1 en } \mathbf{a}: N^{-1}\sum_{i=1}^{N} E\left[\mathbf{x}_i\times g_i^h\left(\mathbf{x}_i'\breve{\mathbf{a}}_N-(\mathbf{x}_i'\tilde{\mathbf{a}}_N)\breve{\boldsymbol{\delta}}_N'\tilde{\mathbf{x}}_i^h,\mathbf{0}\right)\right]=\mathbf{0}\\ \text{CO1 en } \boldsymbol{\delta}: -N^{-1}\sum_{i=1}^{N}\tilde{\mathbf{x}}_i^h\times(\mathbf{x}_i'\tilde{\mathbf{a}}_N)\times g_i^h\left(\mathbf{x}_i'\breve{\mathbf{a}}_N-(\mathbf{x}_i'\tilde{\mathbf{a}}_N)\breve{\boldsymbol{\delta}}_N'\tilde{\mathbf{x}}_i^h,\mathbf{0}\right)=\mathbf{0}\end{array}\right.}$$

- Sous :

$$H_0 : \boldsymbol{\delta}_0 = \mathbf{0}$$

on a :

$$p\lim_{N\to+\infty} \tilde{\mathbf{a}}_N = \mathbf{a}_0 \ \text{ et } \ p\lim_{N\to+\infty} (\breve{\mathbf{a}}_N, \breve{\boldsymbol{\delta}}_N) = (\mathbf{a}_0, \boldsymbol{\delta}_0) = (\mathbf{a}_0, \mathbf{0})$$

et :

$$\begin{cases} p\lim_{N\to+\infty} \ \text{C01 en } \mathbf{a} : E\left[\mathbf{x}_i \times g_i^h(\mathbf{x}_i'\mathbf{a}_0, \mathbf{0})\right] = \mathbf{0} \\ p\lim_{N\to+\infty} \ \text{C01 en } \boldsymbol{\delta} : E\left[-\tilde{\mathbf{x}}_i^h \times (\mathbf{x}_i'\mathbf{a}_0) \times g_i^h(\mathbf{x}_i'\mathbf{a}_0, \mathbf{0})\right] = \mathbf{0} \end{cases},$$

i.e. les équations de score (as.) du Probit « standard » avec :

$$g_i^h(\mathbf{a}, \mathbf{0}) = \varphi(\mathbf{a}'\mathbf{x}_i) \times \frac{y_i - \Phi(\mathbf{a}'\mathbf{x}_i)}{\left(1 - \Phi(\mathbf{a}'\mathbf{x}_i)\right) \times \Phi(\mathbf{a}'\mathbf{x}_i)}.$$

- En outre :

$$p\lim_{N\to+\infty}\Phi\left(\mathbf{x}_i'\breve{\mathbf{a}}_N-(\mathbf{x}_i'\tilde{\mathbf{a}}_N)\breve{\boldsymbol{\delta}}_N'\tilde{\mathbf{x}}_i^h\right)=\Phi(\mathbf{x}_i'\mathbf{a}_0)\ \text{ sous } H_0:\boldsymbol{\delta}_0=\mathbf{0},$$

ce qui implique qu'on peut tester simplement l'hypothèse jointe :

$$H_0:\boldsymbol{\delta}_0=\mathbf{0}$$

en utilisant un test (de Wald) standard après estimation du modèle Probit « avec la variable $(\mathbf{x}_i'\mathbf{a}_0)\times\tilde{\mathbf{x}}_i^h$ ajoutée » :

$$\begin{cases} y_i=1 \text{ si } \tilde{y}_i>0 \\ y_i=0 \text{ si } \tilde{y}_i\leq 0 \end{cases} \text{avec } \tilde{y}_i\,|\,\mathbf{x}_i\sim\mathcal{N}\left(\mathbf{x}_i'\mathbf{a}_0-\boldsymbol{\delta}_0'\tilde{\mathbf{x}}_i^h(\mathbf{x}_i'\mathbf{a}_0),1\right),$$

i.e. sans « correction » de la statistique du test pour l'estimation de $\tilde{\mathbf{a}}_N$ en première étape dans :

$$(\breve{\mathbf{a}}_N,\breve{\boldsymbol{\delta}}_N)\equiv\arg\max_{(\mathbf{a},\boldsymbol{\delta})} N^{-1}\sum_{i=1}^N\left\{\begin{matrix} y_i\times\ln\Phi\left(\mathbf{x}_i'\mathbf{a}-(\mathbf{x}_i'\tilde{\mathbf{a}}_N)\boldsymbol{\delta}'\tilde{\mathbf{x}}_i^h\right) \\ +(1-y_i)\times\ln\left[1-\Phi\left(\mathbf{x}_i'\mathbf{a}-(\mathbf{x}_i'\tilde{\mathbf{a}}_N)\boldsymbol{\delta}'\tilde{\mathbf{x}}_i^h\right)\right]\end{matrix}\right\}.$$

- L'idée sous-jacente ici que si $H_0 : \boldsymbol{\delta}_0 = \mathbf{0}$ n'est pas vérifiée, alors en considérant le programme de maximisation définissant $(\breve{\mathbf{a}}_N, \breve{\boldsymbol{\delta}}_N)$:

$$\max_{(\mathbf{a},\boldsymbol{\delta})} N^{-1} \sum_{i=1}^{N} \left\{ \begin{array}{c} y_i \times \ln \Phi\left(\mathbf{x}_i' \mathbf{a} - (\mathbf{x}_i' \tilde{\mathbf{a}}_N) \boldsymbol{\delta}' \tilde{\mathbf{x}}_i^h \right) \\ +(1 - y_i) \times \ln\left[1 - \Phi\left(\mathbf{x}_i' \mathbf{a} - (\mathbf{x}_i' \tilde{\mathbf{a}}_N) \boldsymbol{\delta}' \tilde{\mathbf{x}}_i^h \right) \right] \end{array} \right\},$$

on a :

$$\begin{cases} p\lim_{N \to +\infty} \text{ C01 en } \mathbf{a} : E\left[\mathbf{x}_i \times g_i^h \left(\mathbf{x}_i' \mathbf{a}_* - (\mathbf{x}_i' \mathbf{a}_*) \boldsymbol{\delta}_*' \tilde{\mathbf{x}}_i^h \right) \right] = \mathbf{0} \\ p\lim_{N \to +\infty} \text{ C01 en } \boldsymbol{\delta} : E\left[-\tilde{\mathbf{x}}_i^h \times (\mathbf{x}_i' \mathbf{a}_*) \times g_i^h \left(\mathbf{x}_i' \mathbf{a}_* - (\mathbf{x}_i' \mathbf{a}_*) \boldsymbol{\delta}_*' \tilde{\mathbf{x}}_i^h \right) \right] = \mathbf{0} \end{cases},$$

avec :

$$\boxed{p\lim_{N \to +\infty} (\breve{\mathbf{a}}_N, \breve{\boldsymbol{\delta}}_N) = (\mathbf{a}_*, \boldsymbol{\delta}_*) \neq (\mathbf{a}_0, \boldsymbol{\delta}_0) \text{ et } \boldsymbol{\delta}_* \neq 0, \text{ en particulier.}}$$

Remarque. ***Spécification, MV et MM*** **(*ou MC*)**

Il est important de noter ici que $P_i^h\left(\mathbf{x}_i'\mathbf{a}_* - (\mathbf{x}_i'\mathbf{a}_*)\boldsymbol{\delta}_*'\tilde{\mathbf{x}}_i^h\right)$ peut être une ***bonne approximation*** de $E[y_i \mid \mathbf{x}_i] = P[y_i = 1 \mid \mathbf{x}_i]$ lorsque le modèle Probit considéré est réellement hétéroscédastique, *i.e*. lorsque $\boldsymbol{\delta}_0 \neq \mathbf{0}$.

Ceci dit, ***on ne peut utiliser les résultats du cadre d'inférence du MV*** puisque (*i*) tous ces résultats reposent sur une spécification correcte du modèle paramétrique estimé et (*ii*) $P[y_i = 1 \mid \mathbf{x}_i] \neq P_i^h\left(\mathbf{x}_i'\mathbf{a}_* - (\mathbf{x}_i'\mathbf{a}_*)\boldsymbol{\delta}_*'\tilde{\mathbf{x}}_i^h\right)$.

Reste alors le cadre d'inférence du Pseudo-MV, voire des MC ou de la MM.

La même remarque s'applique pour le modèle Probit « standard » lorsque $\boldsymbol{\delta}_0 \neq \mathbf{0}$ mais $\boldsymbol{\delta}_0 \simeq \mathbf{0}$, et ce même si $p\lim_{N\to+\infty} \tilde{\mathbf{a}}_N \neq \mathbf{a}_0$ et $p\lim_{N\to+\infty} \tilde{\mathbf{a}}_N \neq \mathbf{a}_*$.

6.3. Spécification des variables latentes (et hétéroscédasticité)

- ***Objectif ici*** :
  - Quelques remarques générales sur le choix de la forme de la variable latente d'un modèle
  - En lien avec la question de l'hétéroscédasticité de son terme d'erreur
  - Montrer :
    - Le caractère restrictif des choix usuels
    - L'intérêt des paramètres aléatoires
    - L'intérêt du non- ou semi-paramétrique

Déjà ± discuté : choix des formes fonctionnelles des modèles « continus »

- Difficile de justifier :

$$\tilde{y}_i \mid \mathbf{x}_i \sim \mathcal{N}\left(\alpha_0 + \tilde{\mathbf{x}}_i'\mathbf{b}_0, \sigma_0^2 \exp(\boldsymbol{\delta}_0'\tilde{\mathbf{x}}_i^h)^2\right)$$

i.e., à la fois :

(*i*) la normalité de $\tilde{y}_i \mid \mathbf{x}_i$

et :

(*ii*) l'hétéroscédasticité de forme exponentielle.

- Ceci explique :

1. certainement pourquoi ce modèle est peu utilisé en pratique.
2. l'intérêt d'autres approches de spécification des modèles à variables latentes

- Trois grands « types » approches alternatives :

  1. ***Approche non-paramétrique*** « stricte »

     Rare en économétrie

  2. ***Approche semi-paramétrique avec index latent***, généralement linéaire

     A la mode en théorie, beaucoup moins en pratique

  3. ***Modèle latent à paramètres aléatoires***

     A la mode en théorie, et en pratique (en choix discret en général)

- Nombreuses autres approches : la « généralisation » des modèles de choix dichotomique est un domaine de recherche très actif …

  - … « le tri se fera » certainement dans quelques années.

Approche non paramétrique stricte

Principe : Estimer la fonction $G_0(\mathbf{x})$ telle que $P[y_i = 1 | \mathbf{x}_i] = G_0(\mathbf{x}_i)$

Avantages : Pas d'hypothèses de formes fonctionnelles ou de distribution

Inconvénients : (Très) « Gros » calculs ; estimations pas pratiques à interpréter ou utiliser

Mon point de vue : très robuste, mais pas assez « applicable » …
… pour l'instant en tourt cas

Approche semi-paramétrique avec index latent linéaire

Principe : Estimer $\boldsymbol{\beta}_0$ dans le modèle $y_i = \mathbf{1}(\mathbf{a}_0'\mathbf{x}_i + u_i > 0)$ avec le minimum d'hypothèses sur $\mathcal{D}(u_i \mid \mathbf{x}_i)$ au-delà d'une hypothèse d'exogénéité.

$$\textit{Normalisation} : \mathbf{a}_0'\mathbf{x}_i = \alpha_0 + \tilde{x}_{1,i} + \boldsymbol{\beta}_0'\tilde{\mathbf{x}}_{(-1),i}$$

Avantages : Peu d'hypothèses de distribution ; paramètres parfois faciles à interpréter

Inconvénients : Gros calculs ou calculs inhabituels ; hypothèses sur $\mathcal{D}(u_i \mid \mathbf{x}_i)$ soit peu usuelles (*e.g.*, travail sur la médiane pour la Méthode du Score Maximum de Manski, 1981), soit assez restrictives (*e.g.* $u_i \perp\!\!\!\perp_{\mathbf{a}_0'\mathbf{x}_i} \mathbf{x}_i$ pour le MV semi-paramétrique de Klein et Spady, 1993) ; paramètres parfois difficiles à interpréter

Mon point de vue : Compromis robustesse/applicabilité peu satisfaisant …
… cumule des désavantages du paramétrique standard et du non- paramétrique

Approche par les paramètres aléatoires

Principe : (i) Modèle de choix : $y_i = \mathbf{1}(\mathbf{a}_i'\mathbf{x}_i + \gamma_0\varepsilon_i > 0)$ avec $\mathcal{D}(\varepsilon_i \mid \mathbf{x}_i) = \mathcal{D}(\varepsilon_i)$ connue, (ii) $\mathcal{D}(\mathbf{a}_i \mid \mathbf{x}_i) = \mathcal{D}(\mathbf{a}_i)$ choisie dans une famille paramétrée par $\boldsymbol{\theta}$. Estimer $\boldsymbol{\theta}_0$ ou $(\boldsymbol{\theta}_0, \gamma_0)$.

Normalisation : $\gamma_0 \equiv 1$ ou $\mathbf{a}_i'\mathbf{x}_i = \alpha_i + \tilde{x}_{1,i} + \boldsymbol{\beta}_i'\tilde{\mathbf{x}}_{(-1),i}$

Avantages : $\tilde{y}_i = \mathbf{a}_i'\mathbf{x}_i + \gamma_0\varepsilon_i$ en tant qu'approximation locale du « vrai modèle » ; hétérogénéité explicite ; facile avec des données de panel ; résultats faciles à interpréter ; le choix de $\mathcal{D}(\varepsilon_i)$ peu problématique (par rapport aux approches alternatives) ; large choix pour $\mathcal{D}(\mathbf{a}_i)$

Inconvénients : Intégration des vraisemblances par simulation ; gros calculs en général ; choix de $\mathcal{D}(\mathbf{a}_i)$ crucial ; semble parfois un peu « trop » facile

Mon point de vue : Un bon compromis robustesse/applicabilité …
… cumule les avantages du paramétrique standard et du non- paramétrique.

A certainement beaucoup bénéficié des efforts de Train pour populariser l'utilisation du MNL à paramètres aléatoires

- ***Intuitions*** :
  - Le modèle $\tilde{y}_{i,t} = \alpha_i + \tilde{\mathbf{x}}'_{i,t}\mathbf{b}_i + \gamma_0 \varepsilon_{i,t}$ est assez peu restrictif. Défini pour chaque *i*. Rend compte explicitement de l'hétérogénéité des préférences, technologies, conditions, …
  - Le choix de la distribution de $\varepsilon_{i,t}$ est d'autant moins important que $\gamma_0 \varepsilon_{i,t}$ est « petit » par rapport à $\alpha_i + \tilde{\mathbf{x}}'_{i,t}\mathbf{b}_i$ (McFadden et Train, 2000)
  - Le choix de $\mathcal{D}(\alpha_i, \mathbf{b}_i)$ est essentiel … mais il est large en pratique, si utilisation correcte des méthodes d'intégration par simulation.

Remarque

Avec :

$$\mathbf{x}_i,\ \mathbf{a}_i \text{ et } \varepsilon_i \text{ mutuellement indépendants}$$

et :

$$\mathbf{a}_0 \equiv E[\mathbf{a}_i],\ \mathbf{\Omega}_0 \equiv V[\mathbf{a}_i] \text{ et, } V[\varepsilon_i] \equiv 1 \text{ et } \gamma_0 \equiv 1$$

on a :

$$\tilde{y}_i = \mathbf{a}_i'\mathbf{x}_i + \varepsilon_i = \mathbf{a}_0'\mathbf{x}_i + u_i \text{ avec } u_i \equiv \varepsilon_i + \mathbf{x}_i'(\mathbf{a}_i - \mathbf{a}_0).$$

et donc :

$$E[u_i \mid \mathbf{x}_i] = 0 \text{ et } V[u_i \mid \mathbf{x}_i] = 1 + \mathbf{x}_i'\mathbf{\Omega}_0\mathbf{x}_i$$

Aussi, si :

$$\mathbf{a}_i \sim \mathcal{N}(\mathbf{a}_0, \mathbf{\Omega}_0) \text{ et } \varepsilon_i \sim \mathcal{N}(0,1)$$

alors :

$$u_i \mid \mathbf{x}_i \sim \mathcal{N}(0, 1 + \mathbf{x}_i'\mathbf{\Omega}_0\mathbf{x}_i)$$

(contrainte de normalisation à imposer) et :

$y_i = \mathbf{1}(\mathbf{a}_i'\mathbf{x}_i + \varepsilon_i > 0)$ suit$|\, \mathbf{x}_i$ un modèle Probit « à hétéroscédasticité quadratique »

- On a :

$$P[y_i = 1 \,|\, \mathbf{x}_i] = \Phi\Big((1 + \mathbf{x}_i'\mathbf{\Omega}_0\mathbf{x}_i)^{-1/2} \times (\mathbf{x}_i'\mathbf{a}_0)\Big).$$

- Forme de fonction de probabilité difficile à utiliser directement dans le cadre du MV(C) en raison :

  (*i*) de l'inverse de la racine carrée

  et :

  (*ii*) de ce que $\mathbf{\Omega}_0$ doit être (semi-)définie positive.

  - L'utilisation d'une décomposition de Cholesky de $\mathbf{\Omega}_0$ résout ces problèmes même si elle ajoute de la non linéarité au modèle.

En général, en pratique on passe (à l'instar de Train ou de Greene) au Logit à paramètres aléatoires