Binary Response Models
Semiparametric Methods in Econometrics
Lecture Notes in Statistics (Springer-Verlag)

Chapter 3 : Binary Response Models

J. L. Horowitz [a]

Avril, 2002

[a]The University of Iowa Iowa City, IA 52242-1000. Joel-horowitz@uiowa.edu

Binary Response Models

Tout le problème tourne autour du modèle binaire :

$$Y = \begin{cases} 1 & if \qquad Y^* > 0 \\ 0 & ifnot \end{cases} \tag{1}$$

$$Y^* = X\beta + U \tag{2}$$

- Si U a une distribution connue (probit, logit)

$\longrightarrow$ estimation par ML

  – **Pb** si la distribution de U est mal spécifiée...

- Si U a une distribution inconnue

  i) et si $U \perp X$ ou U ne dépend que de $X\beta$

  $\longrightarrow$ Single index models

  PB : pas d'hétéroscédasticité ! !

  ii) Si on veut de l'hétéroscédascitié

  $\longrightarrow$ Random Coefficient Models

Petit rappel sur les Single index models *(Chapitre 2)*

Définition : Soit Y v.a et X un $\overrightarrow{v.a}$, $(1 \times k)$; un Single Index Model est un modèle de la forme :

$$\mathbb{E}[Y|x] = G(x\beta) \tag{3}$$

avec β paramètre $(k \times 1)$ inconnu et G fonction inconnue.
Modèle très général :
si $G = Id \longrightarrow$ modèle linéaire
si Y binaire et $G = \Phi$ ou Cdf logistic $\longrightarrow$ modèle probit ou logit.

Utilité des single index, exemples, autres approches... voir book

Identification dans les Single index models

β n'est pas identifié si :

- G est constante
- $\exists \alpha \in \mathbb{R}^k$ et $\exists c \in \mathbb{R}$ tels que $X\alpha = c$ (colinéarité parfaite)

β est identifié à une constante additive (*"location normalisation"*) et multiplicative (*"scale normalisation"*) près.
En effet, soit $G^*(\gamma + \delta\nu) = G(\nu), \forall \nu \in Supp\{X\beta\}$ alors :

$$\mathbb{E}[Y|x] = G(x\beta) \quad (4)$$
$$\mathbb{E}[Y|x] = G^*(\gamma + x\delta\nu) \quad (5)$$

Les modèles (4) et (5) sont indifférenciables empiriquement

Single Index Whith X discrete

(x_1, x_2)	$\mathbb{E}[Y|x]$	$G(x_1 + \beta_2 x_2)$
$(0,0)$	0	$G(0)$
$(1,0)$	0.1	$G(1)$
$(0,1)$	0.3	$G(\beta_2)$
$(1,0)$	0.4	$G(1+\beta 2)$

Tab. 1 – Exemple de Single index model non identifié

En fait, ici, si l'on suppose $G()$ strictement croissante, (on n'a pas d'identification de β) mais on a $\beta_2 > 1$
En outre, si l'on ajoute :

$(0.6, 0.5)$	$.035$	$G(0.6 + 0.5\beta)$

alors on a $\beta_2 < 1.2$

Identification dans les Single index models (suite)

Théorème 2.1 *(page16)*
If
– $\beta_1 = 1$
– *G is differentiable and not constant on Supp*$\{X\beta\}$
– *The components of* X *are continuously distributed , having a joint prob. dist.*
– *Supp*$\{X\} \subsetneq \mathbb{R}^{k-1}$
Then *both* β *and* $G()$ *are identifiable*

Il est possible toutefois d'avoir qq X discrets au prix d'hypothèses supplémentaires.
– *G satisfies a non-periodicity condition*
– *varying values on discrete components do not divide Supp*$\{X\beta\}$ *into disjoint subsets*

Identification avec X_1 continue ; X_2 discrete

Soit $X_1 \in [0,1]$ et $X_2 \in \{0,1\}$ avec $X_1 \perp X_2$ G() supposée croissante et $\beta_1 = 1$
On a $X\beta = X_1 + \beta_2 X_2$

$\mathbb{E}[Y\|x_1, 0]$	$=$	$G(x_1)$		Identification de G sur [0,1]
$\mathbb{E}[Y\|x_1, 1]$	$=$	$G(x_1 + \beta_2)$	Si $\beta_2 > 1$	Supp.$\{x_1 + \beta_2\}$ disjoint de [0,1] : Pas d'ident
			Si $\beta_2 < 1$	Support de $x_1 + \beta_2 \subset [0,1]$ OK

Remarques :

1. Pour l'identification de β if faut que $\mathbb{E}[Y|x]$ soit constant si x bouge de telle sorte que $x\beta$ est constant.
2. G differentiable $\Rightarrow G(X\beta)$ proche de $G(c)$ si $X\beta$ proche de c et donc $\forall c \in Supp\{X\beta\}$; $\{X|0 < d(X\beta, c) < \epsilon\}$ a une proba non nulle.
3. En fait il semble que "X_1 continuously distributed" suffit (p.15).
4. d'une manière générale : si l'on ne suppose rien sur $G()$, on ne peut rien savoir de β

Random Coefficient Model

Idée : Prise en compte des goûts individuels non observables (= random)

$$Y = \begin{cases} 1 & if \quad Y^* > 0 \\ 0 & ifnot \end{cases}$$

$$\begin{aligned} Y^* &= X(\beta + \nu) + V \\ &= X\beta + (X\nu + V) \end{aligned}$$

On retrouve l'équation (1)

$$Y^* = X\beta + U$$

Avec $U = X\nu + V$, le coef de X est initialement $\beta + \nu$ où ν est un $\overrightarrow{va}$

β est la moyenne ou médiane de de la distribution

ν caractérise les dispersions des goûts autour de β

Le modèle admet l'hétéroscédasticité, mais ona deux gros problèmes qui sont l'**identification** et l'**estimation**

Identification de β

On a un truc qui ressemble à un modèle linéaire (sauf Y^*).
Hypothèse de "**scale normalisation**" , i.e. $\beta_1 = 1$
Dans un modèle linéaire classique (Y^* observable) l'hypothèse $\mathbb{E}[U|x] = 0$ "**location normalisation**", serait suffisante pour identifier β, ici non ! (il faut une Hyp. plus forte $U \perp X$ par ex.)
Pourquoi ? Prenons un logit (page 57)
Pour tout $x \in Supp\{X\}$ et pour $b \neq \beta$ on peut construire V_x v.a. telle que

$\mathbb{E}[V_x|x] = 0$

b vérifie les propriétés de "*scale normalisation*"

$F_{V_x}(xb) = F_U(x\beta)$ avec $F()$ CDF,

Théorème 3.2 *(p.59) If*

- $Median[U|x] = 0, \; \forall x \in Supp\{X\}$
- $\beta_1 = 1$

(a) The support of the distribution of X is not contained in any proper linear subspace of $\Re^k$

(b) For almost every $\tilde{x} = (x_2, \cdots, x_k)$, the distribution of $X_1|\tilde{X} = \tilde{x}$ has an everywhere positive density

Qu'est ce que cela veut dire ?
(a) $\Rightarrow$ pas de relation linéaire entre les composantes de X
Remarque :
$P[y = 1|x] = P[U > x\beta|x] \quad (= 1 - P[U \leq x\beta|x] \,)$
et donc (hypothèse sur la médiane)

$$P[y = 1|x] \quad \begin{array}{lll} \geq 0.5 & if & x\beta \geq 0 \\ < 0.5 & if & x\beta < 0 \end{array} \tag{7}$$

(b) Soit $b \neq \beta$, b vérifiant la *"scale normalisation"*
et soit $S_1(b) = \{x\ : x\beta < 0 \leq xb\}$ et $S_2(b) = \{x\ : xb < 0 \leq x\beta\}$
$P[S_1(b) \cup S_2(b)] > 0 \Rightarrow$ identification

Cela veut dire que si il existe un sous-ensemble $\Omega \in Supp\{X\}$ tel que $\forall x \in \Omega P[y = 1|x] < 0.5$ pour β et $P[y = 1|x] \geq 0.5$ pour b ; β identifiable
Comme $\beta_1 = 1$ et $b_1 = 1$ on peut exprimer la condition sur les $S_i(b)$ par la condition (b) ■

Identification et Estimation de $G(x) = P[Y = 1|x]$

Pb :

$P[Y = 1|x]$ n'est pas une fonction continue des composants continus de X
$P[Y = 1|x]$ est connu *via* $x\beta$ et on a (on sait que)

$$P[Y = 1|x] \begin{cases} \geq 0.5 & if \quad x\beta \geq 0 \\ = 0.5 & if \quad x\beta = 0 \\ < 0.5 & if \quad x\beta < 0 \end{cases} \qquad (8)$$

Du coup "$P[Y = 1|x]$ *is not identifiable over most of the support of* X" et on ne peut que des bornes, plus ou moins fines...
Remarque :
Si on suppose $P[Y = 1|x]$ $(= \mathbb{E}[Y\,|X = x])$ est une fonction continue des composants continus de X, alors estimation non-paramétrique classique ..

Estimation de β : Maximum Score Estimator and LAD

Rappelons que

$$\begin{aligned} Median[Y|x] &= inf\{y \ : \ P[Y \geq y|\ x\ ] \geq 0.5\} \\ &= \mathbb{I}\{x\beta \geq 0\} \end{aligned}$$

β $\exists$, ! et minimise le critère de déviation absolue :

$$\beta = Arg \min_{b_1=1} S_{bin}(b)$$

avec

$$\begin{aligned} S_{bin}(b) &= \mathbb{E}|Y - \mathbb{I}\{Xb \geq 0\}| \\ &= \mathbb{E}|Y - (2Y-1)\mathbb{I}\{Xb \geq 0\}| \end{aligned}$$

Théorem 3.3 et 3.4

Théorème 3.3 *:*
If
– $Median[U|x] = 0,\ \forall x \in Supp\{X\}$
– $\beta_1 = 1$
(a) The support of the distribution of X is not contained in any proper linear subspace of $\Re^k$
(b) For almost every $\tilde{x} = (x_2, \cdots, x_k)$, the distribution of $X_1|\tilde{X} = \tilde{x}$ has an everywhere positive density
Then *β is the unique minimizer of $\mathbb{E}[S_{bin}(b)]$*

L'équivalent empirique de $\mathbb{E}[S_{bin}(b)]$ est :

$$\tilde{S}_n(b) = \frac{1}{n}\sum_{i=1}^{n} Y_i - \frac{1}{n}\sum_{i=1}^{n}(2Y_i - 1)\mathbb{I}\{X_i b \geq 0\}$$

Pourquoi “score” ?

$$
\begin{aligned}
Arg \min_{b_1=1} S_{bin}(b) &= Arg \min_{b_1=1} \frac{1}{n}\sum_{i=1}^{n} Y_i - \frac{1}{n}\sum_{i=1}^{n}(2Y_i-1)\mathbb{I}\{X_i b \geq 0\} \\
&= Arg \max_{b_1=1} \frac{1}{n}\sum_{i=1}^{n}(2Y_i-1)\mathbb{I}\{X_i b \geq 0\} \\
&= Arg \max_{b_1=1} \frac{1}{n}\sum_{i=1}^{n}(2Y_i-1)\left(2\cdot\mathbb{I}\{X_i b \geq 0\}-1\right)
\end{aligned}
$$

or il s’agit d’une somme de “scores” (1 si juste, -1 si faux) puisque :

$$
\begin{aligned}
(2Y_i-1)\left(2\cdot\mathbb{I}\{X_i b \geq 0\}-1\right) &= 1 \quad ssi \qquad Y_i = 1 \textbf{ et } X_i\beta \geq 0 \\
&= -1 \quad sinon
\end{aligned}
$$

Théorème 3.4 *:*

If

Hyp de 3.3

$\exists B$ compact set including β

$b_n = Arg \min_{b_1=1\ b\in B} S_{ms}(b)$

Then

$$b_n \xrightarrow{a.s.} \beta \ \ if \ \ ; n \to \infty$$

Pb 1 : Vitesse de convergence $= n^{-1/3}$

Pb 2 : Dist. de $n^{-1/3}\big(b_n-\beta\big)$ est compliquée $\longrightarrow$ Bootstrap

Pb 3 : *“No formal proof that the asymptotic bootstrap distribution of $n^{-1/3}(b_n^*-b_n)$ converges to the asymptotic distribution of $n^{-1/3}(b_n-\beta)$”...*

Smoothed Maximum Score Estimator

Idée : $S_{ms}(b) = \frac{1}{n}\sum_{i=1}^{n}(2Y_i - 1)\mathbb{I}\{X_i b \geq 0\}$ est une fonction discontinue de b...et donc on pourrait la lisser :

$$S_{sms}(b) = \frac{1}{n}\sum_{i=1}^{n}(2Y_i - 1)K\left(\frac{X_i b}{h_n}\right)$$

avec les hypothèses classiques sur K on a

$$K\left(\frac{X_i b}{h_n}\right) \stackrel{n\to\infty}{\longrightarrow} \mathbb{I}\{X_i b \geq 0\}$$

Remarque :
si $X_i b = 0$ ce peut ne pas être vrai, mais $P[X_i b = 0] = 0$

Théorème 3.5 *:*
Sous "certaines" hypothèses, (voir plus loin)
et si *$\{b_n\}$ séquence de solutions de*

$$Maximize\ :\ S_{sms}(b) = \frac{1}{n}\sum_{i=1}^{n}(2Y_i - 1)K\left(\frac{X_i b}{h_n}\right)$$

then

if $nh_n^{2s+1} \stackrel{n\to\infty}{\longrightarrow} \infty$

then $h_n^s(b_n - \beta) \stackrel{P}{\longrightarrow} -Q^{-1}A$

if $nh_n^{2s+1} \stackrel{n\to\infty}{\longrightarrow} \lambda$ *(fini)*

then $(nh_n)^{1/2}(b_n - \beta) \stackrel{D}{\longrightarrow} N\left(-\lambda^{-1/2}Q^{-1}A\ ,\ Q^{-1}DQ^{-1}\right)$

Avec :

$Q = p\lim_{n\to\infty} \frac{\partial^2 S_{sms}(b_n^*)}{\partial b \partial b'}$ (equivalent d'un Hessien)

$D = \int K^2 \cdot \mathbb{E}[X' X p[0|X]]$

et A est abominable...

Les hypothèses

[**1-3**] itou théoreme 3.3 (a), (b), $median(U|x) = 0$, $\beta_1 = 1...+$
$0 < P[Y = 1|x] < 1$ et $\widetilde{\beta} \in$ Compact de $\mathbb{R}^{k-1}$

[**6-9**] K est un noyau d'ordre s

[**5-7**] $X'X$ et $X'XX'X$ admettent des moments finis (existence de Q, D et A)

[**6**] $nh_n \overset{n\to\infty}{\longrightarrow} 0$ et $\frac{log(n)}{nh_n^4} \overset{n\to\infty}{\longrightarrow} 0$

[**10**] Q est negative definite

Remarques :

- Il y a un biais asymptotique à estimer$(-\lambda^{-1/2}Q^{-1}A)$
- En fait il y a des~partout, puisque $\beta_1 = 1$, il s'agit donc des autres composantes des vecteurs matrices, etc...
- Il montre que l'on peut estimer tout ça, le biais, A, Q, D comme il faut..

Autres remarques

⋆ Comme $(nh_n)^{1/2}(\widetilde{b}_n - \widetilde{\beta})$ est normalement distribué, on peut tester des hypothèses sur $\widetilde{\beta}$: on a un t-test et un χ^2, c'est cool.

⋆ *"Under certain conditions"*, le bootstrap donne une meilleure approximation de la distribution que la distribution théorique asymptotique.

⋆ Horowitz donne une expression analytique de la fenêtre optimale pour le $\widetilde{b}_{SMS}$ qui est : $h_{n,opt} = \left(\frac{\lambda^*}{n}\right)^{1/(2s+1)} \longrightarrow$ Plug-in!!

Choice based samples

On peut traiter des échantillons stratifiés suivant les valeurs de Y, i.e. non purement aléatoires.
La proportion d'obs telles que $Y = 1$ est connue et X est aléatoire conditionné par Y

On a alors
n_j = nombre d'obs pour lequel $Y_i = j$ $(n = n_1 + n_2)$
$\pi_1 = P[Y = 1]$ et $\pi_2 = P[Y = 0]$ proportions connues (par ailleurs !!)

Choice based samples

Théorème 3.6 *:*
If
- $Median[U|x] = 0, \ \forall x \in Supp\{X\}$
- $\beta_1 = 1$

$\forall j = 0, 1$ $\{X_i :, i = 1, \cdots, n\}$ *is a random sample from the distribution of* $X|Y = j$
(a) The support of the distribution of X *is not contained in any proper linear subspace of* $\Re^k$
(b) For almost every $\tilde{x} = (x_2, \cdots, x_k)$*, the distribution of* $X_1|\tilde{X} = \tilde{x}$ *has an everywhere positive density*
Then $b_n \overset{a.s.}{\to} \beta$ *as* n_1 *and* $n_2 \to \infty$

avec
n_j = nombre d'obs pour lequel $Y_i = j$ $(n = n_1 + n_2)$
$\pi_1 = P[Y = 1]$ et $\pi_2 = P[Y = 0]$ **connus**

$$
\begin{aligned}
b_n = arg\min S_{n,CB}(b) &= \frac{\pi_1}{n_1} \sum_{i=1;Y_i=1}^{n} \mathbb{I}\{X_i b \geq 0\} \\
&+ \frac{\pi_2}{n_2} \sum_{i=1;Y_i=0}^{n} \mathbb{I}\{X_i b \geq 0\}
\end{aligned}
$$

Panel Data

$$
Y_{it} = \begin{cases} 1 & if \qquad Y_{it}^* > 0 \\ 0 & ifnot \end{cases} \tag{9}
$$

$$
Y_{it}^* = X_{it}\beta + U_i + \epsilon_{it} \tag{10}
$$

En gros c'est assez comparable :
Si $T = 2$ et ϵ_{i1} et ϵ_{i2} ont même distribution $|(X_{i1}, X_{i1})$
Alors

$$
\tilde{Y}_i = \begin{cases} 1 & if \qquad Y_{i2}^* - Y_{i2}^* > 0 \\ 0 & ifnot \end{cases} \tag{11}
$$

et donc on s'en sort avec un estimateur de score....

Panel Data

On a alors l'estimateur équivalent au score, qui maximise :

$$Arg \max_{b_1=1} S_{MS,Pan}(b) = \frac{1}{n}\sum_{i=1}^{n}(Y_{i2} - Y_{i1})\mathbb{I}\{(X_i2 - X_i2)b \geq 0\}$$

Il faut juste que l'évenement $Y_{i2} \neq Y_{i1}$ ait une proba non nulle + quelques autres hypothèses
On peut généraliser la procédure à $T > 2$

Ordered-response Models

Suppose Y prends plus de 2 valeus et que

$$Y = \mathbb{I}\{\alpha_{m-1} < Y^* \leq \alpha_{m-1}\} \ ; \ m = 1, \cdots, M$$

avec

$Y^* = X\beta + U$

Les constantes α_m connues ou inconnues

$Median[U|x] = 0$

On définit alors

$$W = \sum_{m=1}^{M} \mathbb{I}\{Y^* > \alpha_m\}$$

et alors

$$Median(W|x) = \sum_{m=0}^{M} \mathbb{I}\{X\beta > \alpha_m\}$$

Ordered-response Models

Soit

$$S_{n,OR}(b) = \frac{1}{n}\sum_{i=1}^{n} |W_i - \sum_{m=0}^{M} \mathbb{I}\{X_i b \geq \alpha_m\}|$$

Theorem 3.11