3.3 Les modèles de régression à variable dépendante qualitative

3.3.1. La nature des modèles à variable dépendante qualitative

Exemples :

i) Etudier la décision de participation à la force de travail des hommes adultes
→ variable dépendante ne peut prendre que 2 valeurs :

$$\begin{cases} Y = 1 & \text{si individu fait partie de la pop. active} \\ Y = 0 & \text{s'il n'en fait pas partie.} \end{cases}$$

→ variable dépendante binaire / dichotomique.
(dpt ntmt du tx de chô, du tx de salaire, ...)

ii) Analyser les déterminants des votes aux USA
Si deux partis : Démocrates et Républicains
→ variable dépendante ne peut prendre que 2 valeurs :

$$\begin{cases} Y = 1 & \text{si vote en faveur des Démocrates} \\ Y = 0 & \text{si vote en faveur des Républicains} \end{cases}$$

(dpt ntmt du tx de croiss. du PIB, des tx de chô et d'inflat° si candidat brigue un nouveau mandat)

! variable dép. peut être multicatégorielle
... on débute par le cas dichotomique

- Différence ds modèles de régression où variable dépendante Y est quantitative & qualitative :

  > Objectif lorsque Y est quantitative = estimer sa valeur moyenne ou espérée, compte tenu des valeurs des régresseurs
  > $\longrightarrow$ estimer $E(Y_i \mid X_{1i}, X_{2i}, \ldots, X_{ki})$

  > Objectif lorsque Y est qualitative = estimer la probabilité d'arrivée d'un événement (p.ex. voter pour Démocrate, posséder une maison, appartenir à un syndicat, pratiquer un sport) $\longrightarrow$ modèles avec réponse souvent appelés des modèles probabilistes.

- Trois approches pour développer un modèle de probabilité à variable dépendante binaire :

  1. Le modèle de probabilité linéaire (MPL)
  2. Le modèle logit
  3. Le modèle probit

3.3.2. Le modèle de probabilité linéaire

- Supposons le modèle de régression suivant :

$$\boxed{Y_i = \beta_1 + \beta_2 X_i + u_i} \qquad (15.2.1)$$

où $\begin{cases} Y = 1 & \text{si famille possède une maison,} \\ Y = 0 & \text{sinon} \end{cases}$ X = revenu familial

Comme variable dpdte binaire / dichotomique, on parle de modèle de probabilité linéaire (MPL).

Ceci s'explique par le fait que l'espérance cond. de Y_i étant donné X_i, càd. $E(Y_i / X_i)$, peut être interprétée comme la prob. cond. que l'événement se produise compte tenu de X_i, càd. $Pr(Y_i = 1 / X_i)$.

Dans notre exemple, $E(Y_i/X_i)$ donne prob. qu'une famille soit propriétaire d'une maison compte tenu de son revenu X_i.

- Explication alternative :

Supposons que $E(u_i) = 0$ (hyp. hab. pour obtenir estimateur non biaisé)

$\Rightarrow$

$$\boxed{E(Y_i / X_i) = \beta_1 + \beta_2 X_i} \qquad (15.2.2)$$

Si P_i = probabilité que $Y_i = 1$ (événemt se produit)
et $(1-P_i)$ = probabilité que $Y_i = 0$ (événemt ne se produit pas)
alors la variable Y_i présente la distribution de prob. suivante

Y_i	Probabilité
0	$1-P_i$
1	P_i
Total	1

$\Rightarrow$ Y_i suit une distribut° de Bernouilli

Par définition, on obtient que :

$$\boxed{E(Y_i) = 0\ (1-P_i) + 1\ (P_i) = P_i} \qquad (15.2.3)$$

Comme expressions (15.2.2) et (15.2.3) sont équivalentes, on trouve que :

$$\boxed{E(Y_i/X_i) = \beta_1 + \beta_2 X_i = P_i} \qquad (15.2.4)$$

Cela signifie que :
- l'espérance cond. de notre modèle (15.2.1) peut être interprétée comme la prob. cond. de Y_i
- l'éspérance d'une variable aléatoire de Bernouilli correspond à la prob. que la var. aléat. soit égale à 1

Rappel :

Une variable aléatoire X suit une distrib. de Bernouilli si sa fction de densité de probabilité (FDP) est la suivante :

$$\boxed{\begin{aligned} P(X=0) &= 1-p \\ P(X=1) &= p \end{aligned}}$$

où p ($\in [0,1]$) est la prob. qu'un événement soit un „succès" (comme p. ex. prob. d'obtenir une face sur une pièce de monnaie)

Pour une telle variable, on a que :

$$\boxed{\begin{aligned} E(X) &= [1 \,.\, P(X=1) + 0 \,.\, P(X=1)] = p \\ \text{var}(X) &= p \cdot q \end{aligned}}$$

où $q = 1-p$
q est la prob. qu'un événement soit un „échec".

Puisque P_i doit se situer lel 0 et 1, on peut poser la restriction suivante :

$$\boxed{0 \leq E(Y_i / X_i) \leq 1} \qquad (15.2.5)$$

Peut-on appliquer les MCO aux MPL ?

Pas si simple, MPL pose tieurs problèmes.

A) L'absence de normalité des erreurs u_i

Hyp. de normalité du terme d'erreur pas soutenable pour MPL car, comme la variable dpdte Y_i, les erreurs u_i ne prennent que 2 valeurs. Terme d'erreur suit églmt une dist. de prob. de Bernouilli.

Pq ?

En réécrivant notre modèle de départ (15.2.1) comme suit :

$$\boxed{u_i = Y_i - \beta_1 - \beta_2 X_i} \qquad (15.2.6)$$

on voit que dist. de prob. du terme d'erreur u_i est la suivante :

	u_i	Probabilité
Lorsque $Y_i = 1$	$1 - \beta_1 - \beta_2 X_i$	P_i
Lorsque $Y_i = 0$	$-\beta_1 - \beta_2 X_i$	$(1 - P_i)$

Non respect de l'hypothèse de normalité pas nécessairement aussi grave qu'il n'y parait.

Pq ?

1) Estimations ponctuelles des paramètres pop. inconnus par MCO sont non biaisées $\hat{m}$ si terme d'erreur ne suit pas une distribution normale ... mais maigre consolation.

2) La distribution des estimateurs des MCO tend générlmt vers une distribution normale lorsque la taille de l'échantillon converge vers l'infini. Par consqt, dans les gds échantillons, on pourra utiliser la technique des MCO pour estimer les MPL et faire de l'inférence statistique.

B) Les variances hétéroscédastiques des erreurs

Même si $E(u_i) = 0$ et $cov(u_i, u_j) = 0$ pour $i \neq j$
$\Rightarrow$ dans MPL, les erreurs sont hétéroscédastiques.

Pq ?

Une var. aléat. suivant distrib. de Bernoulli a une moyenne et une variance théoriques qui valent resp. p et $p(1-p)$, où p = prob. de succès.

Terme d'erreur ds MPL suit distr. de Bernoulli

$$\Rightarrow \quad \boxed{var(u_i) = P_i(1-P_i)} \qquad (15.2.7)$$

Comme $P_i = E(Y_i / X_i) = \beta_1 + \beta_2 X_i$, on voit que var (u_i) dépend des valeurs prises par X
$\Rightarrow$ terme d'erreur hétérosc.

En présence d'hétérosc., estimateurs des MCO (bien que non biaisés) ne sont pas efficients.
Cependant, pblme pas tjs insurmontable : $\exists \neq$ solut° potent.

Comme var (u_i) dépend de $E(Y_i / X_i)$, on peut transformer modèle de régression (15.2.1) en le divisant par :

$$\sqrt{E(Y_i / X_i)\ [1 - E(Y_i / X_i)]} = \sqrt{P_i\,(1 - P_i)} = \sqrt{w_i} \qquad (15.2.8)$$

$\Rightarrow$

$$\boxed{\frac{Y_i}{\sqrt{w_i}} = \frac{\beta_1}{\sqrt{w_i}} + \beta_2 \cdot \frac{X_i}{\sqrt{w_i}} + \frac{u_i}{\sqrt{w_i}}} \qquad (15.2.9)$$

$\equiv$ Méthode des moindres carrés pondérés (MCP)
où w_i = poids

Pblme : en pratique la vraie $E(Y_i / X_i)$ est inconnue et donc les poids aussi.

$\Rightarrow$ On estime les w_i par procédure en 2 étapes :

1) On estime modèle initial (15.2.1) par MCO et on obtient $\hat{Y}_i$ (càd. estimat° de $E(Y_i / X_i)$)
Ensuite, on calcule $\hat{w}_i$ par l'expression suivante : $\hat{w}_i = \hat{Y}_i\,(1 - \hat{Y}_i)$ ($\hat{w}_i$ = estimat° de w_i).

2) On utilise les $\hat{\omega}_i$ pour transformer les données, et ensuite on estime éq. transformée (15.2.9) par MCO.

Cependant, si n suffisant gd, possibilité de calculer erreurs standards de White (1980) robustes à l'hétéros. asympt.

c) <u>La non satisfaction de $0 \leq E(Y_i/X_i) \leq 1$</u>

$E(Y_i/X_i)$ peut ne pas être compris lel 0 et 1.

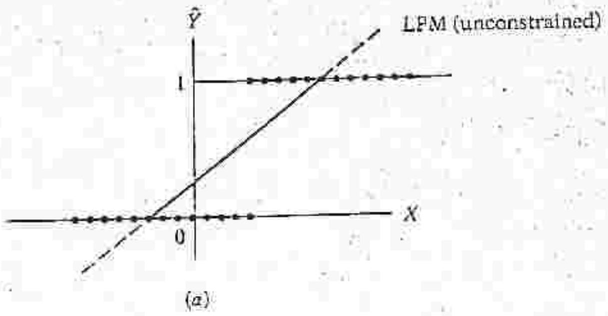
<u>Intuition</u> : ds MPL, $E(Y_i/X_i)$ mesure prob. cond. de l'occurence d'un événement Y étant donné X. Bien que cond. vraie a priori, rien ne garantit que $\hat{Y}_i$ satisferont cette condition.

<u>Deux moyens pour déterminer si $\hat{Y}_i \in [0,1]$</u> :

1) Estimer MPL par MCO et vérifier si cond. satisfaite.
   Si certains $\hat{Y}_i < 0$, on leur attribue valeur 0.
   Si certains $\hat{Y}_i > 1$, on leur attribue valeur 1.
2) Imaginer procédure d'estimat° qui garantisse que prob. cond. estimées ($\hat{Y}_i$) soient comprises lel 0 et 1.
   (cf. modèles logit et probit)

D). La valeur contestable du R^2 comme mesure de la qualité d'un ajustement

R^2 habituel est peu adéquat.



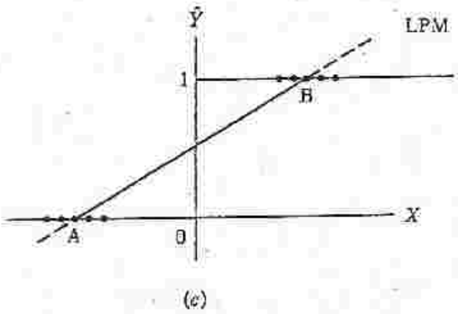


FIGURE 15.1 Linear probability models.

Dans plupart des applications, R^2 faible (sut lel 0,2 et 0,6). D'autres statistiques, tq pseudo - R^2, sut préférées au R^2 traditionnel pour mesurer qualité d'ajustmt avec MPL.

3.3.3. Applications du MPL

Exemple numérique :

TABLE 15.1
HYPOTHETICAL DATA ON HOME OWNERSHIP ($Y = 1$ IF OWNS HOME, 0 OTHERWISE) AND INCOME X (THOUSANDS OF DOLLARS)

Family	Y	X	Family	Y	X
1	0	8	21	1	22
2	1	16	22	1	16
3	1	18	23	0	12
4	0	11	24	0	11
5	0	12	25	1	16
6	1	19	26	0	11
7	1	20	27	1	20
8	0	13	28	1	18
9	0	9	29	0	11
10	0	10	30	0	10
11	1	17	31	1	17
12	1	18	32	0	13
13	0	14	33	1	21
14	1	20	34	1	20
15	0	6	35	0	11
16	1	19	36	0	8
17	1	16	37	1	17
18	0	10	38	1	16
19	0	8	39	0	7
20	1	18	40	1	17

A partir des données, on estime MPL par MCO :

$$\hat{Y}_i = -0,9457 + 0,1021\ X_i \qquad (15.2.10)$$

$$\begin{array}{lll} (0,1228) & (0,0082) & \longrightarrow \text{s.e} \\ (-7,6984) & (12,515) & \longrightarrow t\text{-stat} \end{array}$$

$R^2 = 0,8048$

Interprétation :

i) Valeur ordonnée à l'origine $(-0,9457)$: probabilité qu'1 famille ayant revenu nul possède une maison.

ii) Valeur du coeff. de pente $(0,1021)$: pour une variation unitaire du revenu (1000 \$), prob. moyenne de posséder une maison augmente de 10% environ.

En effet, $\beta_2 = \dfrac{\Delta P}{\Delta X}$ ⇒ pour obtenir variat° relative en % de la prob. suite à variat° unitaire de X : $\beta_2 * 100$.

iii) On peut estimer prob. de posséder maison p. ex pour $X = 12$:

$$(\hat{Y}_i \mid X = 12) = -0{,}9457 + 12\,(0{,}1021) = 0{,}2795.$$

TABLE 15.2
ACTUAL Y, ESTIMATED Y, AND WEIGHTS w_i FOR THE HOME OWNERSHIP EXAMPLE

Y_i	$\hat{Y}_i$	$\hat{w}_i$‡	$\sqrt{\hat{w}_i}$	X_i	Y_i	$\hat{Y}_i$	$\hat{w}_i$‡	$\sqrt{\hat{w}_i}$	X_i
0	−0.129*			8	1	1.301†			22
1	0.688	0.2148	0.4633	16	1	0.688	0.2147	0.4633	16
1	0.893	0.0956	0.3091	18	0	0.280	0.2016	0.4990	12
0	0.178	0.1463	0.3825	11	0	0.178	0.1463	0.3825	11
0	0.280	0.2016	0.4490	12	1	0.688	0.2147	0.4633	16
1	0.995	0.00498	0.0705	19	0	0.178	0.1463	0.3825	11
1	1.098†			20	1	1.097†			20
0	0.382	0.2361	0.4859	13	1	0.893	0.0956	0.3091	18
0	−0.0265*			9	0	0.178	0.1463	0.3825	11
0	0.076	0.0702	0.2650	10	0	0.076	0.0702	0.2650	10
1	0.791	0.1653	0.4066	17	1	0.791	0.1653	0.4055	17
1	0.893	0.0956	0.3091	18	0	0.382	0.2361	0.4859	13
0	0.484	0.2497	0.4997	14	1	1.199†			21
1	1.097†			20	1	1.097†			20
0	−0.333*			6	0	0.178	0.1463	0.3825	11
1	0.995	0.00498	0.0705	19	0	−0.129*			8
1	0.688	0.2147	0.4633	16	1	0.791	0.1653	0.4066	17
0	0.076	0.0702	0.2650	10	1	0.688	0.2147	0.4633	16
0	−0.129*			8	0	−0.231*			7
1	0.893	0.0956	0.3091	18	1	0.791	0.1653	0.4066	17

*Treated as zero to avoid probabilities being negative.
†Treated as unity to avoid probabilities exceeding one.
‡$\hat{Y}_i(1 - \hat{Y}_i)$.

- Présence de 6 valeurs estimées $(\hat{Y}_i) < 0$ et $6 > 1$
  $\Rightarrow$ $\hat{m}$ si $E(Y_i/X_i)$ nécess. comprise ht 0 et 1 (car peut être interprétée comme prob. cond.), les $\hat{Y}_i$ peuvent être < 0 ou > 1.

- Autre pblme : hétéroscédasticité
  $\Rightarrow$ on ne peut pas se baser sur erreurs standards estimées (15.2.10) pour faire inf. stat.
  $\Rightarrow$ on peut utiliser MCP pour obtenir estimateurs plus efficients

$$\hat{w}_i = \hat{Y}_i\,(1 - \hat{Y}_i)$$

$$\boxed{\hat{w}_i = \hat{Y}_i\,(1-\hat{Y}_i)}$$

- Lorsque $\hat{Y}_i < 0$ ou $\hat{Y}_i > 1 \Rightarrow \hat{w}_i$ prend valeur nég.
Comme $\sqrt{}$ d'un chiffre nég. $\nexists$, on ne peut utiliser les obs. associées aux pondérations nég. dans estim. par MCP $\Rightarrow$ # d'obs. chute de 40 à 28.
Pour éviter de perdre des degrés de liberté, l'alternative consiste à supposer que $\begin{cases} \hat{Y}_i = 0{,}01 & \text{si } \hat{Y}_i < 0 \\ \hat{Y}_i = 0{,}99 & \text{si } \hat{Y}_i > 1 \end{cases}$

- <u>En appliquant MCP aux 28 obs</u> :

$$\frac{\hat{Y}_i}{\sqrt{w_i}} = \underset{\substack{(0{,}1206) \\ (-10{,}332)}}{-1{,}2456}\ \frac{1}{\sqrt{w_i}} + \underset{\substack{(0{,}0069) \\ (17{,}454)}}{0{,}1196}\ \frac{X_i}{\sqrt{w_i}} \qquad (15.2.11)$$

$\longrightarrow$ s.e.

$\longrightarrow$ t-stat

$R^2 = 0{,}9214$

- <u>Par rapport aux résultats par MCO</u> :
erreurs standards + faibles $\longrightarrow |t\text{-stat}|$ + gdes

Mais prudence car :

1) 12 obs. supprimées $\longrightarrow$ estimation moins précise.
2) poids w_i sont estimés $\longrightarrow$ procédures hab. de test d'hyp sont, au sens strict, uniq. valables pour gds éch.
3) pblme de normalité du terme d'erreur demeure sans doute car échant. de petite taille.

• Autre application du MPL : étude de Cohen et al. (1970)

Analyse de participat° à la force de travail de ≠ cat. de salariés en fction de var. socio-éco et démographiques.

TABLE 15.3 LABOR-FORCE PARTICIPATION
Regression of women, age 22 and over, living in largest 96 standard metropolitan statistical areas (SMSA) (dependent variable: in or out of labor force during 1966)

Explanatory variable	Coefficient	t ratio
Constant	0.4368	15.4
Marital status		
Married, spouse present	—	—
Married, other	0.1523	13.8
Never married	0.2915	22.0
Age		
22–54	—	—
55–64	−0.0594	−5.7
65 and over	−0.2753	−9.0
Years of schooling		
0–4	—	—
5–8	0.1255	5.8
9–11	0.1704	7.9
12–15	0.2231	10.6
16 and over	0.3061	13.3
Unemployment rate (1966), %		
Under 2.5	—	—
2.5–3.4	−0.0213	−1.6
3.5–4.0	−0.0269	−2.0
4.1–5.0	−0.0291	−2.2
5.1 and over	−0.0311	−2.4
Employment change (1965–1966), %		
Under 3.5	—	—
3.5–6.49	0.0301	3.2
6.5 and over	0.0529	5.1
Relative employment opportunities, %		
Under 62	—	—
62–73.9	0.0381	3.2
74 and over	0.0571	3.2
FILOW, $		
Less than 1,500 and negative	—	—
1,500–7,499	−0.1451	−15.4
7,500 and over	−0.2455	−24.4
Interaction (marital status and age)		
Marital status / Age		
Other / 55–64	−0.0406	−2.1
Other / 65 and over	−0.1391	−7.4
Never married / 55–64	−0.1104	−3.3
Never married / 65 and over	−0.2045	−6.4
Interaction (age and years of schooling completed)		
Age / Years of schooling		
65 and over / 5–8	−0.0885	−2.8
65 and over / 9–11	−0.0848	−2.4
65 and over / 12–15	−0.1288	−4.0
65 and over / 16 and over	−0.1628	−3.6
	$R^2 = 0.175$	
No. of observations = 25,153		

Note: — indicates the base or omitted category.
FILOW: family income less own wage and salary income.
Source: Malcolm S. Cohen, Samuel A. Rea, Jr., and Robert I. Lerman, *A Micro Model of Labor Supply*, BLS Staff Paper 4, U.S. Department of Labor, 1970, Table F-6, pp. 212–213.

$Y = 1$ si personne fait partie de la pop. active.

$Y = 0$ si personne inactive

Coeff. de pente = tx de variat° (ou variat° relative) de la prob. cond. d'arrivée d'un évènmt suite à variat° unitaire de la valeur d'une variable expl.

- <u>Variable dummy „ 65 ans et plus "</u> :

  coeff = − 0,2753 ⇒ ceteris paribus, prob. de participat° à la force de travail des ♀ de 65 ans et plus est en moyenne inférieure d'environ 27% à celle de la cat. de base (càd ♀ de 22 à 54 ans).

- <u>Variable dummy „ 16 ans de scolarité ou plus"</u> :

  coeff = 0,3061 ⇒ ceteris paribus, prob. que les ♀ avec 16 ans de scolarité ou plus participent à la force de travail est en moyenne d'environ 31% supérieure à celle de la cat. de réf. (càd. ♀ avec moins de 5 ans d'éducat°).

- <u>Terme d'interaction relatif au statut marital et à l'âge</u>:

  Prob. de participat° à force de travail :
  i) supérieur de ± 29% pour femmes jamais mariées // cat. de réf. (càd. ♀ mariée dont époux présent).
  ii) plus faible de ± 28% pour femmes de 65 ans et + // cat. de réf. (càd. ♀ de 22 à 54 ans).

  Quid des ♀ „ jamais mariées" et „ 65 ans et plus" ?

  $$
  \begin{aligned}
  (Y_i / X_i) &= +\,0,2915 \ (\text{dummy „ jamais mariée"}) \\
  &\quad -\,0,2753 \ (\text{dummy „ 65 ans et +"}) \\
  &\quad -\,0,2045 \ (\text{dummy „ jamais mariée" } \underline{\&} \text{ „ 65 +"}) \\
  &= -\,0,1883
  \end{aligned}
  $$

  ⇒ Femmes jamais mariées de 65 ans et + ont une prob. d'être active de ± 19% infér. à cat. réf. (♀ de 22-54 ans, mariée époux prés.

- Femmes de 65 ans et +, jamais mariées, participent + à la force de travail que celles qui ont 65 ans et + qui sont mariées ou appartiennent à la catég. autre.

Pq ?

Prob de participer au marché du travail (// à la catég. de réf, càd ♀ mariées avec époux présent et âgées de 22 à 54 ans) est de :

* 28 % inférieure pour ♀ de 65 ans et plus, mariées

$$\left[(\hat{Y}_i / X_i) = 0 \text{ (dummy „mariée")} \right.$$
$$- 0,2753 \text{ (dummy „65 ans et +")}$$
$$\left. = - 0,2753 \right]$$

* 26 % inférieure pour ♀ de 65 ans et plus, dans cat. „mariée autre"

$$\left[(\hat{Y}_i / X_i) = + 0,1523 \text{ (dummy „mariée autre")} \right.$$
$$- 0,2753 \text{ (dummy „65 ans ou +")}$$
$$- 0,1391 \text{ (dummy „mariée autre et 65 ans et +")}$$
$$\left. = - 0,2621 \right]$$

- Possibilité d'estimer prob. cond. de participat° au marché du travail de ≠ cat. de personnes.
Exemple, pour les ♀ „mariées autre", âgées de 22 à 54 ans, ayant 12 à 15 ans de scolarité, avec tx de chô de 2,5 à 3,4% (ds la région), une variat° de l'emploi de 3,5 à 6,9% et des opportunités d'emploi de 74% et plus, la prob est de ± 88%.

3.3.4. Alternatives au MPL

- MPL connaît plusieurs problèmes :
  i) Non - normalité du terme d'erreur u_i,
  ii) Hétéroscédasticité du terme d'erreur u_i,
  iii) Possibilité que $\hat{Y}_i \notin [0, 1]$,
  iv) Valeurs du R^2 svt faibles (indicateur peu adéquat).

- Mais pblmes pas tjs insurmontables :
  Hétérosc $\longrightarrow$ MCP
  Non - normalité $\longrightarrow$ accroître taille de l'éch.
  $\hat{Y}_i \notin [0, 1]$ $\longrightarrow$ recourir aux techniques de progr. math. pour garantir que prob. cond. estimées soient comprises lel 0 et 1.

- Cpdt, problme fondamental avec MPL :
  Prob. cond. d'un „succès" ($P_i = E(Y_i = 1 / X_i)$ dépend linéairement de la variable X.
  Effet marginal de X sur la var. dép. est tjs constant. (quelle que soit valeur initiale de X)

  Dans exemple // possession d'une maison :
  Lorsque X (le revenu) ↑ d'une unité (1000 \$), prob. de posséder une maison ↑ tjs de 0,10 (10%) quel que soit le niveau de revenu initial.

  ↳ Caractéristique peu conforme à la réalité
  Anticipat°: prob. cond. ne dpd pas linéairmt de X_i

On s'attend à ce que :

i) Pour très faibles revenus, une famille ne possède pas de maison.

ii) Pour niveau de revenu suffisamment élevé, X^*, elle soit probablmt propriétaire.

iii) Croiss. du revenu au-delà de X^* ait peu d'effet sur probabilité de posséder une maison.

⇒ A chaque extrémité de la distrib. des revenus, prob. d'être propriétaire d'une maison (très) peu sensible à une variation de X.

⇒ Besoin d'un modèle (probabiliste) ayant caract suivantes :

a) lorsque $X \uparrow$, prob. cond. $\uparrow$ ($P_i = E(Y_i = 1 / X_i)$) mais ne sort jamais de l'intervalle 0 - 1.

b) relation lel P_i et X_i non linéaire, càd. P_i converge vers 0 de moins en moins vite lorsque X devient très petit et P_i tend vers 1 de moins en moins vite lorsque X devient très grand.

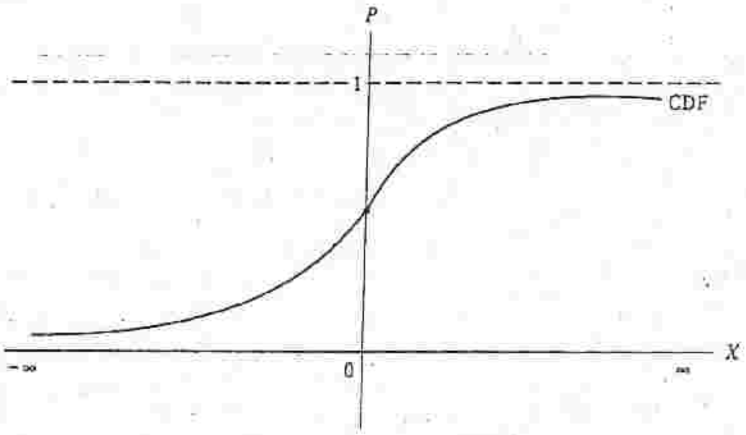


FIGURE 15.2 A cumulative distribution function (CDF).

Caractéristiques :

i) probabilité $\in [0, 1]$.

ii) probabilité ne dpd pas linéairmt de X_i.

iii) courbe en S (sigmoïde) ressemble bcp à la FDC d'une variable aléatoire (p.ex. # de fautes de frappe à l'ordinateur, taille humaine)

(Rappel : FDC d'une var.-aléat. X = prob. qu'elle prenne valeur inférieure ou égale à x_0, où x_0 est une valeur numér. spécifiée de X. $F(X=x_0) = P(X \le x_0)$.

⇒ on peut utiliser une FDC pour modèles de régr. où var. dép. est dichotomique, prenant valeurs 0 et 1.

⇒ Quelle FDC utiliser ?

(~~Une~~ seule FDC par var. aléat.)

Pour raisons hist. et pratiques, on utilise génerlmt :
i) fction logistique ⟶ modèle logit
ii) fction normale ⟶ modèle probit (ou normit)

3.3.5. Le modèle logit

Exemple : prob. de posséder une maison

$$\boxed{MPL \equiv P_i = E(Y_i = 1 / X_i) = \beta_1 + \beta_2 X_i} \qquad (15.5.1)$$

où X_i = revenu de la famille i.
$Y_i = 1$ si famille i possède une maison.

Considérons représentat° alternat° de possession d'une maison :

$$\boxed{P_i = E(Y_i / X_i) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}}} \qquad (15.5.2)$$

$\Rightarrow$

$$\boxed{P_i = E(Y_i / X_i) = \frac{1}{1 + e^{-Z_i}} = \frac{e^{Z}}{1 + e^{Z}}} \qquad (15.5.3)$$

où $Z_i = \beta_1 + \beta_2 X_i$

Fct° de distrib. logistique cumulative

On peut démontrer que :
i) Lorsque $Z_i \in [-\infty, +\infty]$, $P_i \in [0, 1]$
ii) P_i ne dpd pas linéairmt de Z_i (de X_i)

Problème :
P_i est non linéaire en X mais aussi ds les paramètres β.
⟶ MCO n'est plus adaptée.

- Pour contourner ce pblme, on linéarise le modèle :

Si prob. posséder une maison $P_i = \dfrac{e^{Z_i}}{1+e^{Z_i}}$, alors

prob. de ne pas en posséder une $1 - P_i = \dfrac{1}{1+e^{Z_i}}$ (15.5.4)

$$\Rightarrow \boxed{\frac{P_i}{1-P_i} = e^{Z_i}} \qquad (15.5.5)$$

$\dfrac{P_i}{1-P_i}$ = ratio de chances („ odds ratio")

en faveur de la possession d'une maison

≡ ratio tel la prob. de posséder une maison et la prob. de ne pas posséder une maison.

Si $P_i = 0,8$ et $1 - P_i = 0,2$ $\longrightarrow$ $\dfrac{P_i}{1-P_i} = 4$

cela signifie que les chances sont de 4 contre 1 en faveur de la possession d'une maison

$$\boxed{L_i = \ln\left(\frac{P_i}{1-P_i}\right) = Z_i = \beta_1 + \beta_2 X_i} \qquad (15.5.6)$$

L ≡ log du ratio de chances ≡ logit

$\longrightarrow$ linéaire ds variable X mais aussi ds para. β

$$\boxed{L_i = \ln\left(\frac{P_i}{1-P_i}\right) = Z_i = \beta_1 + \beta_2 X_i}$$

Caractéristique des modèles logit :

i) Lorsque P se situe lel 0 et 1, càd lorsque Z varie lel $-\infty$ et $+\infty$, L s'étend lel $-\infty$ et $+\infty$.
→ Bien que prob. $P \in [0, 1]$, les logits L ne sont pas bornés de cette façon.

ii) Bien que L linéaire en X, P n'est pas linéaire en X ($\neq$ imp. avec MPL).

Dans modèle logit : $\boxed{dP/dX = \beta_2 P(1-P)}$

→ tx de variat° de la prob. // X inclut non slmt β_2 mais aussi prob. à partir de laquelle variat° est mesurée.

→ l'effet d'une variat° unitaire de X sur P est + fort lorsque $P = 0,5$ et moindre lorsque P proche de 0 ou de 1.

Exemple :
si $\beta_2 = 1$ et $P = 0,5$ → $\frac{dP}{dX} = 1 . 0,5 . 0,5 = 0,25$
si $\beta_2 = 1$ et $P = 0,9$ → $\frac{dP}{dX} = 1 . 0,9 . 0,1 = 0,09$

iii) Modèle logit peut inclure autant de régresseurs que nous souhaitons.

iv) Interprétation du modèle logit ?

- Coeff. de pente β_2 :
  - mesure la variat° de L suite à une variation unitaire de X $(\beta_2 = \Delta L / \Delta X)$.
  - indique la variat° du log des chances en faveur de la possession d'une maison lorsque le revenu change d'une unité (p. ex. 1000 \$).

- Valeur de l'intercepte β_1 :
  - indique valeur du log des chances de posséder une maison lorsque X (le revenu) est nul.
  - cette interprétat° peut n'avoir aucune signif. concrète.

v) Calcul de la prob. de posséder une maison pour valeur particulière de X (X^*) :

$P_i = \dfrac{1}{1+e^{-Z}}$ peut être utilisé à cet effet, une fois disponibles les estimations de β_1 et β_2.

vi) MPL : prob. cond. P_i dépend linéairmt de X_i
Modèle logit : (log du) ratio des chances linéairmt relié à X_i.

3.3.6. L'estimation du modèle logit

Modèle logit $\equiv$ $\boxed{L_i = \ln\left(\dfrac{P_i}{1-P_i}\right) = \beta_1 + \beta_2 X_i}$ (15.6.1)

Estimation dépend du type de données :

i) individuelles ou micro,

ii) groupées ou répétées.

A) Les données au niveau individuel

Avec données individuelles (cf. exemple // à la possession d'une maiso estimat° d'un modèle logit (15.6.1) par MCO est impossible.

Pq ?

$P_i = 1$ si famille possède une maison
$P_i = 0$ si elle n'en possède pas

En imputant ces valeurs dans le logit L_i :

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) = \ln\left(\frac{1}{0}\right) \quad \text{si famille possède maison}$$

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) = \ln\left(\frac{0}{1}\right) \quad \text{si famille n'en possède pas}$$

Les expressions n'ont pas de sens car :

$$\left.\begin{array}{l} \ln\left(\frac{1}{0}\right) = \ln(\infty) = \infty \\ \ln\left(\frac{0}{1}\right) = \ln(0) \quad \nexists \end{array}\right\} \text{logit } L_i \text{ prend valeur } \infty \text{ ou n'existe pas}$$

$\Rightarrow$ Estimat° d'un modèle logit est impossible avec MCO lorsque données indiv. ou micro-économét.

$\Rightarrow$ Utilisation de la méthode du maximum de vraisemblance

Estimation par MV (max de vraisemblance) du modèle logit contenant des données individuelles (non groupées)

- *Exemple* : But = estimer prob. qu'un individu possède une maison compte tenu de son revenu X.

- *Hyp.* : cette probabilité peut être exprimée par une fction logistique tq :

$$\boxed{P_i = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}}} \qquad (1)$$

On observe pas P_i mais uniq. le résultat $\longrightarrow$ on observe que $Y_i = 1$ si individu possède une maison et $Y_i = 0$ sinon.

Y_i suit une distribution de Bernouilli :

$$P(Y_i = 1) = P_i \qquad (2)$$

$$P(Y_i = 0) = 1 - P_i \qquad (3)$$

- *Supposons que* :
  i) L'on dispose d'un échant. aléat. de n observ.
  ii) $f_i(Y_i)$ = la probabilité que $Y_i = 0$ ou $Y_i = 1$

  Remarque : $f_i(Y_i) = P_i$ si $Y_i = 1$
  $f_i(Y_i) = 1 - P_i$ si $Y_i = 0$

$\Rightarrow$ probabilité jointe d'observer les n valeurs de Y,

$$\boxed{f(Y_1, Y_2, \dots, Y_n) = \prod_1^n f_i(Y_i) = \prod_1^n P_i^{Y_i} (1 - P_i)^{1 - Y_i}} \qquad (4)$$

Rem : • si $Y_i = 1 \Rightarrow P_i^{Y_i} (1 - P_i)^{1 - Y_i} = P_i^1 (1 - P_i)^{1-1} = P_i$
si $Y_i = 0 \Rightarrow P_i^{Y_i} (1 - P_i)^{1 - Y_i} = P_i^0 (1 - P_i)^{1-0} = 1 - P_i$

$f(Y_1, Y_2, \ldots, Y_n)$ = $\underline{\text{fction de vraisemblance}}$ (FV)
(likelihood function, LF)

- En prenant log naturel de (4) :

$$\ln f(Y_1, Y_2, \ldots, Y_n) = \sum_1^n [Y_i \ln P_i + (1-Y_i) \ln (1-P_i)]$$
$$= \sum_1^n [Y_i \ln P_i - Y_i \ln (1-P_i) + \ln (1-P_i)]$$
$$= \sum_1^n \left[Y_i \ln \left(\frac{P_i}{1-P_i}\right)\right] + \sum_1^n \ln (1-P_i) \quad (5)$$

- Nous savons que :

$$P_i = \frac{1}{1+e^{-(\beta_1+\beta_2 X_i)}}$$

$$(1-P_i) = \frac{1}{1+e^{\beta_1+\beta_2 X_i}} \quad (6)$$

$$\ln \left(\frac{P_i}{1-P_i}\right) = \beta_1 + \beta_2 X_i \quad (7)$$

- En remplaçant (6) et (7) dans (5) :

$$\boxed{\ln f(Y_1, Y_2, \ldots, Y_n) = \sum_1^n Y_i (\beta_1+\beta_2 X_i) - \sum_1^n \ln [1 + e^{(\beta_1+\beta_2 X_i)}]} \quad (8)$$

Pq ? (2ème terme de (8))

$$+\sum \ln (1-P_i) = +\sum \ln\left(\frac{1}{1+e^{z}}\right) = +\sum \ln(1) - \ln (1+e^{z})$$
$$= +\sum \ln (1) - \sum \ln (1+e^{z}) = -\sum \ln (1+e^{z})$$

car $\ln (1) = 0$

$\ln f(Y_1, Y_2, \ldots, Y_n) \equiv$ fction log de vraisemblance (FLV)
(log likelihood function, LLF)

FLV dépend des paramètres β_1 et β_2 (X_i et Y_i connus)

Méthode d'estimation par max de vraisemblance → objectif = maximiser FLV, càd. obtenir valeur pour paramètres inconnus β_1 et β_2 qui max. prob. d'observer les Y_i (donnés)

⇒ On différencie FLV (expression (8)) // aux inconnues β_1 et β_2 ; on annule les expressions qui en résultent et on résout ces dernières.

Expressions qui en résultent non linéaires dans les paramètres (aucune solut° explicite) → pour obtenir solution chiffrée recours à méthode d'estimation non linéaire.

Lorsque valeurs numériques obtenues pour β_1 et β_2, calcul prob. cond. de posséder maison à partir de :

$$\boxed{P_i = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}}}$$

Remarque :

Méthode du max. de vraisemblance pour le modèle probit est identique à celle du modèle logit, mais comme expression de la prob. cond. (expression (1)) on utilise la FDC normale plutôt que logistique. Expression qui en résulte assez compliquée mais idée générale identique.

B) Les données groupées ou répétées

Pour chaque niveau de revenu X_i, il y a N_i familles parmi lesquelles n_i sont propriétaires de leur maison ($n_i \leq N_i$)

TABLE 15.4 HYPOTHETICAL DATA ON X_i (INCOME), N_i (NUMBER OF FAMILIES AT INCOME X_i), AND n_i (NUMBER OF FAMILIES OWNING A HOUSE)

X (thousands of dollars)	N_i	n_i
6	40	8
8	50	12
10	60	18
13	80	28
15	100	45
20	70	36
25	65	39
30	50	33
35	40	30
40	25	20

[21]For a comparatively simple discussion of maximum likelihood in the context of the logit model, see John Aldrich and Forrest Nelson, op. cit., pp. 49–54. See also, Alfred Demarsi, *Logit Modeling: Practical Applications*, Sage Publications, Newbury Park, Calif., 1992.

[22]From elementary statistics recall that the probability of an event is the limit of the relative frequency as the sample size becomes infinitely large.

Nous pouvons calculer :

$$\boxed{\hat{P}_i = \frac{n_i}{N_i}} \qquad (15.6.2)$$

où $\hat{P}_i$ = <u>fréquence relative</u> qui peut être utilisée comme estimation de P_i (càd. prob. cond. de posséder une maison étant donné X_i)

Si N_i assez gd (càd. assez de familles par classe de revenu) alors $\hat{P}_i$ bonne estimation de P_i (car prob. d'un évènmt = limite de la fréquence relative lorsque taille de l'éch. devient infiniment gde).

- Logit estimé s'écrit comme suit :

$$\boxed{\hat{L}_i = \ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right) = \hat{\beta}_1 + \hat{\beta}_2 X_i} \qquad (15.6.3)$$

$\hat{L}_i$ fournit bonne estimat° de L_i si N_i assez gd pour chque X_i
Avec données groupées, valeurs de la var. dpdte (càd. du logit) put être estimées.

- Peut-on estimer modèle logit (15.6.3) par MCO ?

Non !

Pq ?

On peut montrer que si N_i est assez gd et si chaque observat° ds une catégorie donnée de revenu X_i est distribuée indpdamment comme une variable binomiale, alors :

$$\boxed{u_i \sim N\left[0\ ,\ \frac{1}{N_i P_i(1-P_i)}\right]} \qquad (15.6.4)$$

Terme d'erreur hétéroscédastique car sa variance cond. dépend de P_i qui lui-m̂ dépend de X_i.

$\Longrightarrow$ Moindres carrés pondérés (MCP)

En pratique, on remplace l'inconnue P_i par $\hat{P}_i$, et on utilise l'expression suivante :

$$\boxed{\hat{\sigma}^2 = \frac{1}{N_i \hat{P}_i(1-\hat{P}_i)}} \text{ comme estimateur de } \sigma^2. \qquad (15.6.5)$$

- Différentes étapes de l'estimation d'un modèle logit (avec des données groupées) :

1) Pour chaque niveau de revenu X_i, calcul de la prob. de posséder une maison :

$$\hat{P}_i = \frac{n_i}{N_i}$$

2) Pour chaque X_i, estimation de la valeur du logit :

$$\hat{L}_i = \ln\left[\frac{\hat{P}_i}{(1-\hat{P}_i)}\right]$$

3) Pour résoudre pblme d'hétérosc., on transforme modèle logit initial (càd. $L_i = \ln\left(\frac{P_i}{1-P_i}\right) = \beta_1 + \beta_2 X_i$) comme suit :

$$\boxed{\sqrt{w_i}\,.\,L_i = \beta_1 . \sqrt{w_i} + \beta_2 . \sqrt{w_i}\, X_i + \sqrt{w_i}\,.\,u_i}$$

$\Rightarrow$ (15.6.6)

$$\boxed{L_i^* = \beta_1 \sqrt{w_i} + \beta_2 . X_i^* + v_i} \qquad (15.6.7)$$

où poids $w_i = N_i \hat{P}_i (1-\hat{P}_i)$

L_i^*, X_i^*, v_i sont les variables L_i, X_i, u_i pondérées/transformées

Remarque :

Nveau terme d'erreur est homoscéd. car :

$$E(u_i)^2 = \sigma^2 / w_i \quad \text{et} \quad v_i = u_i \sqrt{w_i}$$

$$\Rightarrow E(v_i)^2 = E(u_i\sqrt{w_i})^2 = w_i\, E(u_i)^2 = w_i . \frac{\sigma^2}{w_i} = \sigma^2.$$

4) On estime l'expression : $L_i^* = \beta_1 . \sqrt{w_i} + \beta_2 . X_i^* + v_i$.
par MCO (régression passant par l'origine).

5) On établit les intervalles de confiance et on teste les hypothèses selon procédure standard avec MCO.
Cpdt, conclusions (à strictmt parler) valables que si l'échantillon est assez grand.

3.3.7. Le modèle logit avec des données groupées (glogit) : un exemple numérique

. Données relatives à la propriété des ménages par catég. de revenu.

TABLE 15.5 DATA TO ESTIMATE THE LOGIT MODEL OF OWNERSHIP

X (thousands of dollars) (1)	N_i (2)	n_i (3)	$\hat{P}_i$ (4) = (3) ÷ (2)	$1-\hat{P}_i$ (5)	$\frac{\hat{P}_i}{1-\hat{P}_i}$ (6)	$\hat{L}_i = \ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right)$ (7)	$N_i\hat{P}_i(1-\hat{P}_i) = w_i$ (8)	$\sqrt{w_i} = \sqrt{N_i\hat{P}_i(1-\hat{P}_i)}$ (9) = $\sqrt{(8)}$	$\hat{L}_i^* = \hat{L}_i\sqrt{w_i}$ (10) = (7)(9)	$X_i^* = X_i\sqrt{w_i}$ (11) = (1)(9)
6	40	8	0.20	0.80	0.25	−1.3863	6.40	2.5298	−3.5071	15.1788
8	50	12	0.24	0.76	0.32	−1.1526	9.12	3.0199	−3.4807	24.1592
10	60	18	0.30	0.70	0.43	−0.8472	12.60	3.5496	−3.0072	35.4960
13	80	28	0.35	0.65	0.54	−0.6190	18.20	4.2661	−2.6407	55.4593
15	100	45	0.45	0.55	0.82	−0.2007	24.75	4.9749	−0.9985	74.6235
20	70	36	0.51	0.49	1.04	0.0400	17.49	4.1825	0.1673	83.6506
25	65	39	0.60	0.40	1.50	0.4054	15.60	3.9497	1.6012	98.7425
30	50	33	0.66	0.34	1.94	0.6633	11.20	3.3496	2.2218	100.4880
35	40	30	0.75	0.25	3.0	1.0986	7.50	2.7386	3.0086	95.8405
40	25	20	0.80	0.20	4.0	1.3863	4.00	2.000	2.7726	80.0000

Résultats du modèle glogit estimé par MCP :

$$\hat{L}_i^* = -1,59474 . \sqrt{w_i} + 0,07862 . X_i^* \qquad (15.7\text{-}1)$$

$$e.s = (0,11046) \qquad (0,00539)$$

$$t = (-14,43619) \qquad (14,56675)$$

$$R^2 = 0,9642$$

Rem : R^2 = coeff. de corrél. au carré lel L_i^* réels (càd. calculés à partir des $\hat{P}_i$) et estimés (à partir du modèle glogit)

A) L'interprétation du modèle logit estimé

Rappel : $\hat{L}^*_i = -1,59474 . \sqrt{w_i} + 0,07862 . X^*_i$ (15.7.1)

$(-14,43619)$ $(14,56675)$ ← t-stat

1. L'interprétation en terme de logit

Coeff. de pente estimé suggère que pour une croissance d'une unité (1000 \$) du revenu (pondéré), le log (pondéré) des chances de posséder une maison augmente de 0,08 unité.

→ interprétat° mécanique, peu attrayante.

2. L'interprétation en termes de chances

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) \Longrightarrow \exp(L_i) = \frac{P_i}{1-P_i} \equiv \text{ratio des chance}$$

Si on prend l'exponentielle de nos résultats (15.7.1) :

$$e^{\sqrt{w_i}} \cdot \left(\frac{\hat{P}_i}{1-\hat{P}_i}\right) = e^{-1,59474 . \sqrt{w_i} + 0,07862 \, X^*_i}$$

(car $\hat{L}^*_i = \sqrt{w_i} . \hat{L}$
$\hat{L}_i = \hat{P}_i/(1-\hat{P}_i$

$$= e^{-1,59474 . \sqrt{w_i}} . e^{0,07862 \, X^*_i} \quad (15.7.2)$$

$$e^{0,07862} = 1,0817$$

Interprétat° : pour croiss. d'une unité du revenu (pondéré), les chances (pondérées) de posséder une maison croissent de 1,0817, càd d'environ 8,17 %.

Règle générale : en prenant exponentielle du j-ème coeff. de pente, qu'on soustrait valeur 1 et qu'on multiplie résultat par 100, on obtient le % de variat° ds les chances pour une croissance

3. Le calcul des probabilités

Langage du logit et du ratio de chance peut paraître peu familier

→ calculer la probabilité de posséder une maison pour un revenu donné.

Supposons que $X = 20$ (20 000 \$), en remplaçant dans résultats de notre régression (15.7.1) :

$$\hat{L}_i^* = -1{,}59474 \sqrt{w_i} + 0{,}07862\, X_i^*$$
$$= -1{,}59474 \sqrt{w_i} + 0{,}07862\, X_i \sqrt{w_i}$$

Comme pour $X_i = 20$, $\sqrt{w_i} = 4{,}1825$ (cf. Tabl. 15.5) :

$$\hat{L}_i^* = -1{,}59474 \cdot 4{,}1825 + 0{,}07862 \cdot 20 \cdot 4{,}1825$$
$$= -0{,}0934$$

Pour obtenir $\hat{L}_i$: diviser ts les termes par $\sqrt{w_i}$ $(= 4{,}1825)$

$$\Rightarrow \hat{L}_i = -0{,}0223$$

$$\Rightarrow \text{pour } X = 20 : \quad \ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right) = -0{,}0223$$

$$\Rightarrow \left(\frac{\hat{P}_i}{1-\hat{P}_i}\right) = e^{-0{,}0223} = 0{,}9779$$

$$\Rightarrow \hat{P}_i = \frac{e^{-0{,}0223}}{1 + e^{-0{,}0223}} = 0{,}4944$$

Pour un revenu de 20000 \$, la prob. qu'une famille possède une maison est d'environ 49%.

TABLE 15.6 LSTAR, XSTAR, ESTIMATED LSTAR, PROBABILITY, AND CHANGE IN PROBABILITY*

X	Lstar	Xstar	ELstar	Logit	Probability, $\hat{P}$	Change in probability†
6	−3.50710	15.1788	−2.84096	−1.12299	0.24545	0.01456
8	−3.48070	24.15920	−2.91648	−0.96575	0.27572	0.01570
10	−3.48070	35.49600	−2.86988	−0.80850	0.30821	0.01676
13	−2.64070	55.45930	−2.44293	−0.57263	0.36063	0.01813
15	−0.99850	74.52350	−2.06652	−0.41538	0.39762	0.01883
20	0.16730	83.65060	−0.09311	−0.02226	0.49443	0.01965
25	1.60120	98.74250	1.46472	0.37984	0.59166	0.01899
30	2.22118	100.48800	2.55896	0.76396	0.68221	0.01704
35	3.00860	95.84050	3.16794	1.15677	0.76074	0.01431
40	2.77260	80.00000	3.10038	1.55019	0.82494	0.01135

*Lstar and Xstar are from Table 15.5. ELstar is the estimated Lstar. Logit is the unweighted logit. Probability is the estimated probability of owning a house. Change in probability is the change per unit change in income.
†Computed from $\hat{\beta}_2 \hat{P}(1 - \hat{P}) = 0.07862 \hat{P}(1 - \hat{P})$.

Probabilités conditionnelles de posséder une maison pour ≠ niveaux de revenus.

Probabilité de posséder une maison augmente avec le revenu mais pas linéairement (comme avec MPL)

4. Le calcul du taux de variation de la probabilité

- Taux de variat° de la probabilité (cf. slide 15.21) :

$$\boxed{\frac{dP_i}{dX_i} = \beta_2 (1 - P_i) P_i}$$

- Variation de la prob. de posséder une maison lorsque X ↗ d'une unité et que X = 20 ?

$$\frac{d\hat{P}_i}{dX_i} = \hat{\beta}_2 (1 - \hat{P}_i) \hat{P}_i$$

Comme $\hat{\beta}_2 = 0{,}07862$ et $\hat{P}_i = 0{,}4944$ (pour $X = 20$) :

$$\frac{d\hat{P}_i}{dX_i} = 0{,}07862 \cdot (1 - 0{,}4944) \cdot 0{,}4944 = 0{,}01965$$

⇒ Variation du revenu de 1000 \$, lorsque le revenu est de 20000 \$, augmente prob. de posséder une maison de 1,97 pts. de %.

- Analyse graphique :

Variation de la prob. de posséder une maison pour ≠ nux de revenus, prend la forme d'une cloche.

Faible variat° de la prob. lorsque X petit ou grand

Lorsque X = 5 : +1,46%
X = 40 : +1,13%

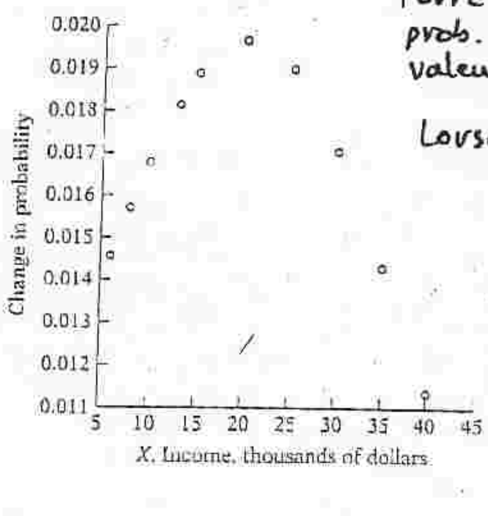


FIGURE 15.3 Change in probability in relation to income.

Forte variat° de la prob. lorsque X a valeur moyenne.

Lorsque X = 20 : +1,97%

3.3.8. Le modèle logit pour des données non groupées ou individuelles

TABLE 15.7 DATA ON THE EFFECT OF PERSONALIZED SYSTEM OF INSTRUCTION (PSI) ON COURSE GRADES

Observation	GPA grade	TUCE grade	PSI	Grade	Letter grade	Observation	GPA grade	TUCE grade	PSI	Grade	Letter grade
1	2.66	20	0	0	C	17	2.75	25	0	0	C
2	2.89	22	0	0	B	18	2.83	19	0	0	C
3	3.28	24	0	0	B	19	3.12	23	1	0	B
4	2.92	12	0	0	B	20	3.16	25	1	1	A
5	4.00	21	0	1	A	21	2.06	22	1	0	C
6	2.86	17	0	0	B	22	3.62	28	1	1	A
7	2.76	17	0	0	B	23	2.89	14	1	0	C
8	2.87	21	0	0	B	24	3.51	26	1	0	B
9	3.03	25	0	0	C	25	3.54	24	1	1	A
10	3.92	29	0	1	A	26	2.83	27	1	1	A
11	2.63	20	0	0	C	27	3.39	17	1	1	A
12	3.32	23	0	0	B	28	2.67	24	1	0	B
13	3.57	23	0	0	B	29	3.65	21	1	1	A
14	3.26	25	0	1	A	30	4.00	23	1	1	A
15	3.53	26	0	0	B	31	3.10	21	1	0	C
16	2.74	19	0	0	B	32	2.39	19	1	1	A

Notes: Grade $Y = 1$ if the final grade is A
$= 0$ if the final grade is B or C
TUCE = score on an examination given at the beginning of the term to test entering knowledge of macroeconomics
PSI = 1 if the new teaching method is used
= 0 otherwise
GPA = the entering grade point average
Source: L. Spector and M. Mazzeo, "Probit Analysis and Economic Education," *Journal of Economic Education*, vol. 11, 1980, pp. 37–44.

- Variable dépendante Y :

„Grade" ≡ „Note" finale d'un étudiant ds un cours de niveau moyen en microéconomie.

$Y = 1$ si note finale de l'étudiant était A

$Y = 0$ si note finale de l'étudiant était B ou C.

- Variables explicatives X :

GPA → note moyenne de l'étudiant lors de son examen d'entrée.

TUCE → score de l'étudiant à un examen au début du trimestre en macro-économie.

PSI → $= 1$ si nouvelle méthode d'enseigmt.
$= 0$ sinon.

- Modèle logit correspondant :

$$\boxed{L_i = \ln\left(\frac{P_i}{1-P_i}\right) = \beta_1 + \beta_2\ GPA + \beta_3\ TUCE + \beta_4\ PSI + u_i} \quad (15.8.1)$$

$P_i = 1$ si note finale de l'étudiant est A (càd. si $Y = 1$)

$P_i = 0$ si note finale étudiant est B ou C (càd. si $Y = 0$)

(P_i = prob. cond. que $Y = 1$ étant donné X)

- Si on introduit directmt P_i dans logit :

$$L_i = \ln\left(\frac{1}{1-1}\right) = \ln\left(\frac{1}{0}\right) = \ln(\infty) = \infty \quad \text{si } Y = 1.$$

$L_i \nexists$ si $Y = 0$ car $\ln\left(\frac{0}{1-0}\right) = \ln(0)$ qui $\nexists$

$\Rightarrow$ ces expression n'ont pas de sens

$\Rightarrow$ MCO ou MCP ne peuvent être utilisés

$\Rightarrow$ recours à procédures d'estimation non linéaires utilisant la méthode du max. de vraisemblance.

- Résultats économétriques :

Plupart des logiciels économétriques ont des sous-programmes permettant d'estimer des modèles logit avec des données individuelles.

TABLE 15.8 REGRESSION RESULTS OF (15.8.1)

Dependent Variable: Grade
Method: ML-Binary Logit
Convergence achieved after 5 iterations

Variable	Coefficient	Std. error	Z statistic	Probability
C	-13.0213	4.931	-2.6405	0.0082
GPA	2.8261	1.2629	2.2377	0.0252
TUCE	0.0951	0.1415	0.67223	0.5014
PSI	2.3786	1.0645	2.2345	0.0255

McFadden $R^2 = 0.3740$ LR statistic (3 df) = 15.40419

• Quelques observations

1) Echantillon doit être de gde taille afin d'interpréter les résultats.

2) On utilise statistique standard normale Z pour évaluer la significativité d'un coefficient de régression.
Pas surprenant car lorsque taille de l'échantillon ↑, la distribut° de Student converge vers la distrib. normale.

3) La mesure classique de la qualité d'un ajustmt (le R^2) n'est pas très intéressante ds modèles à var. dép. binaires.

⇒ ∃ ᐩieurs mesures alternatives au R^2 traditionnel : " pseudo R^2 "

a) Le R^2 de McFadden $\in [0,1]$

$$\boxed{R^2_{McF} = 1 - (FLV_{nr} / FLV_r)}$$

où

FLV_{nr} : fct° log de vraisemblance non restreinte, càd où tous les régresseurs sont inclus ds modèle.

FLV_r : fct° log de vraisemblance restreinte, càd où seule la valeur en ordonnée à l'origine figure dans le modèle.

Conceptuellmt, $FLV_{nr} \equiv SCR$ (Σ des carrés résid.)
$FLV_r \equiv SCT$ (Σ des carrés totale).

Dans notre exemple, $R^2_{McF} = 0{,}3740$: pas mal.

b) Le „R^2 de dénombrement" $\in [0,1]$

$$\boxed{R^2 \text{ de dénombrement} = \frac{\text{nbre de prévisions correctes}}{\text{nbre total d'observations}}}$$

Fonctionnement ?

Comme var. dép. d'un modèle logit $= 0$ ou 1, si prob estimée est : $> 0,5$ $\longrightarrow$ on lui attribue la valeur 1,

$< 0,5$ $\longrightarrow$ on lui attribue la valeur 0.

Ensuite, on dénombre le nbre de prévisions (d'estimations) correctes et on calcule le R^2 de dénombrement.

4) Pour tester hyp. nulle que ts coeff. de pente sont simultanément égaux à zéro, on utilise la „statistique du ratio de vraisemblance" (RV) ("likelihood ratio statistic" LR)

$\Rightarrow$ équivalent du F-test ds régression linéaire.

Sous H_0 (que ts coeff de pente nuls), statistique du RV suit distributᵒ de χ^2 dont nbre degrés de liberté = nbre de variables expl. (ds notre exemple, $dl = 3$).

exclure intercepte ds calcul des dl

Interprétation des résultats

TABLE 15.8 REGRESSION RESULTS OF (15.8.1)
Dependent Variable: Grade
Method: ML-Binary Logit
Convergence achieved after 5 iterations

Variable	Coefficient	Std. error	Z statistic	Probability
C	-13.0213	4.931	-2.6405	0.0082
GPA	2.8261	1.2629	2.2377	0.0252
TUCE	0.0951	0.1415	0.67223	0.5014
PSI	2.3786	1.0645	2.2345	0.0255

McFadden $R^2 = 0.3740$ LR statistic (3 df) = 15.40419

$$\boxed{L_i = \ln\left(\frac{P_i}{1-P_i}\right) = \beta_1 + \beta_2\, GPA + \beta_3\, TUCE + \beta_4\, PSI + u_i}$$

- Chaque coeff de pente = coeff. de pente partielle $\rightarrow$ mesure la variation du logit estimé suite à un chgmt unitaire dans la valeur de chaque régresseur (les autres cts).

  Coeff. var. GPA signifie que ceteris paribus lorsque GPA ↑ d'une unité, logit estimé augmente en moyenne de 2,83 unités.

  Relat° positive aussi lel $\hat{L}_i$ et autres régresseurs mais non sign. pour var. TUCE.

  RV = 15,40 (p-value = 0,0015) $\rightarrow$ RH₀ (rejet de l'hyp. selon laquelle coeff. de pente ts nuls).

- Interprétation en termes de chances : prendre exp. ≠ coeff. pente.

  Ex : coeff PSI = 2,3786 (PSI = 1 si nulle méth. d'enseig.
  = 0 sinon)

  exp (2,3786) = 10,7897

  Signif. ? Etudiants soumis à nulle méthode d'enseig. ont 10x plus de chances d'obtenir un A que ceux qui n'y sont pas soumis (ceteris paribus).

- Calcul de la probabilité effective qu'un étudiant obtienne un A :

Prenons p. ex. 10ème étudiant de notre éch. (cf. Table 15.7, slide 15.35) $\longrightarrow$ on introduit données de l'étudiant de modèle logit estimé (cf. Table 15.8)

$\Rightarrow$

$$\hat{L}_i = \ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right) = \hat{\beta}_1 + \hat{\beta}_2 . GPA_i + \hat{\beta}_3 \; TUCE_i + \hat{\beta}_4 \; PSI_i$$

$$= 13{,}0213 + 2{,}8261\ (3{,}92) + 0{,}0951\ (29) + 2{,}3786\ (0)$$

$$= 0{,}8149$$

$$\Rightarrow \frac{\hat{P}_i}{1-\hat{P}_i} = e^{0{,}8149} = 2{,}2589$$

$$\Rightarrow \hat{P}_i = (1-\hat{P}_i) . 2{,}2589$$

$$\Rightarrow \hat{P}_i = 2{,}2589 - 2{,}2589 . \hat{P}_i$$

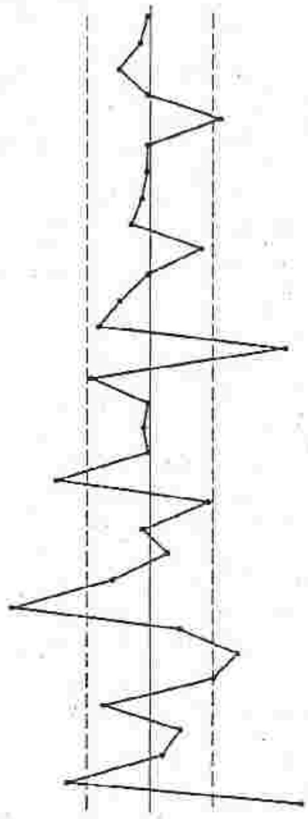
$$\Rightarrow (1 + 2{,}2589)\ \hat{P}_i = 2{,}2589$$

$$\Rightarrow \hat{P}_i = \frac{2{,}2589}{3{,}2589} = 0{,}69$$

Probabilité que le 10ème étudiant obtienne une note A, étant donné ses caract., est de 69%.

TABLE 15.9 ACTUAL AND FITTED VALUES BASED ON REGRESSION IN TABLE 15.8

Observation	Actual	Fitted	Residual	Residual plot
1	0	0.02658	−0.02658	
2	0	0.05950	−0.05950	
3	0	0.18726	−0.18726	
4	0	0.02590	−0.02590	
5	1	0.56989	0.43011	
6	0	0.03486	−0.03486	
7	0	0.02650	−0.02650	
8	0	0.05156	−0.05156	
9	0	0.11113	−0.11113	
10	1	0.69351	0.30649	
11	0	0.02447	−0.02447	
12	0	0.19000	−0.19000	
13	0	0.32224	−0.32224	
*14	1	0.19321	0.80679	
15	0	0.36099	−0.36099	
16	0	0.03018	−0.03018	
17	0	0.05363	−0.05363	
18	0	0.03859	−0.03859	
*19	0	0.58987	−0.58987	
20	1	0.66079	0.33921	
21	0	0.06138	−0.06138	
22	1	0.90485	0.09515	
23	0	0.24177	−0.24177	
*24	0	0.85209	−0.85209	
25	1	0.83829	0.16171	
*26	1	0.48113	0.51887	
27	1	0.63542	0.36458	
28	0	0.30722	−0.30722	
29	1	0.84170	0.15830	
30	1	0.94534	0.05466	
*31	0	0.52912	−0.52912	
*32	1	0.11103	0.88897	



*Incorrect predictions.

Tableau présente valeurs réelles (observées) et estimées (ou prévues / ajustées) de la var. dép. pour notre exemple.

Sur 32 observations : 6 prévisions incorrectes
(rappel : si prob. estimée $> 0,5 \rightarrow 1$; $< 0,5 \rightarrow 0$)

R^2 de dénombrement $= \frac{26}{32} = 0,8125$

R^2 de McFadden $= 0,3740$

Deux mesures pas directement comparables mais fournissent une idée de la qualité de l'ajustement ... mais il ne faut pas surestimer l'imp. de la qualité de l'ajust. lorsque

3.3.9. Le modèle probit

Pour expliquer comptmt d'une variable dpdte dichotomique, il faut utiliser une FDC appropriée.

Modèle logit → FDC logistique :

$$P_i = E(Y_i / X_i) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}}$$

Modèle probit (normit) ⟶ FDC normale :

$$P_i = E(Y_i / X_i) = \int_{-\infty}^{\beta_1 + \beta_2 X_i} \frac{1}{\sqrt{2\pi}} \cdot e^{-s^2/2} \, ds$$

et ensuite m̂ dévelpmt que pour logit.

<u>Rappel</u> : Si variable Y suit une dist. normale, de moyenne μ et de variance σ^2 :

→ FDP : $f(Y) = \frac{1}{\sqrt{2\sigma^2 \pi}} \cdot e^{-(Y-\mu)^2/2\sigma^2}$

→ FDC : $f(Y) = \int_{-\infty}^{X_0} \frac{1}{\sqrt{2\sigma^2 \pi}} \cdot e^{-(Y-\mu)^2/2\sigma^2}$

Au lieu de suivre cette voie, présentat° du modèle probit sur base de la <u>théorie de l'utilité de McFadden (1973)</u>.

Exemple : possession d'une maison par une famille.

Hyp 1 : décision de la i-ème famille de posséder une maison dépend d'un „indice d'utilité I_i" inobservable („la variable latente")

Hyp 2 : indice d'utilité est déterminé par une ou plusieurs variables explicatives (p. ex. le revenu X_i)

Hyp 3 : lorsque valeur de I_i ↑, prob. que i-ème famille possède une maison ↑.

$$\boxed{I_i = \beta_1 + \beta_2 X_i} \qquad (15.9.1)$$

où X_i est le revenu de la i-ème famille.

Comment indice I_i (inobservable) est-il relié à la décision réelle d'acheter une maison ?

Soit $Y = 1$ si famille possède une maison.
$Y = 0$ si famille n'en possède pas.

Raisonnable de faire hyp. qu'il existe un „seuil" ou un „niveau critique" de l'indice d'utilité, I_i^*, tq si :
$I_i \geq I_i^*$ famille achète maison
$I_i < I_i^*$ famille n'en achète pas.

Seuil I_i^* et indice I_i ne sont pas observables, mais si on suppose que les 2 variables suivent une distribution normale avec les m̂ moyennes et variances, il est possible d'estimer paramètres de l'indice d'utilité (15.9.1) mais aussi d'obtenir info sur l'indice d'utilité lui-m̂.

Comment procéder ?

Étant donné l'hypothèse de normalité, la probilité que le seuil I_i^* soit inférieur ou égal à l'indice d'utilité I_i peut être calculée à partir de la FDC normale standardisée (ou standard) :

$$\boxed{\begin{aligned} P_i = P(Y_i = 1 / X_i) &= P(I_i^* \leq I_i) \\ &= P(Z_i \leq \beta_1 + \beta_2 X_i) \\ &= F(\beta_1 + \beta_2 X_i) \end{aligned}} \qquad (15.9.2)$$

où $P(Y=1/X)$ indique prob. qu'une famille possède une maison étant donné son revenu et Z_i est une variable suivant une distributᵒ normale standard, càd $Z \sim N(0,1)$.

- Si on remplace ds l'Éq. (15.9.2) la fct° F (FDC d'une variable normale standardisée) par son expression explicite :

$$\boxed{F(\beta_1+\beta_2 X_i) = F(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-Z^2/2}\, dZ = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1+\beta_2 X_i} e^{-Z^2/2}\, dZ} \quad (15.9.3)$$

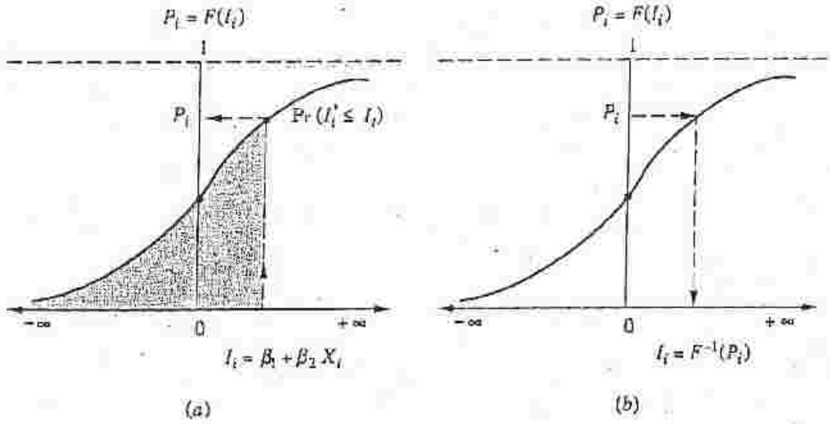


FIGURE 15.4 Probit model: (a) given I_i, read P_i from the ordinate; (b) given P_i, read I_i from the abscissa.

Probabilité de posséder une maison (P) mesurée par la surface de la courbe standard normale allant de $-\infty$ à I_i (cf. Figure (a))

Pour obtenir info sur valeur de l'indice I_i (ainsi que sur valeur des paramètres β_1 et β_2) $\Rightarrow$ prendre inverse de la prob. de posséder une maison P_i (càd. 15.9.2) :

$$\boxed{I_i = F^{-1}(P_i) = F^{-1}[F(I_i)] = F^{-1}[F(\beta_1+\beta_2 X_i)] = \beta_1+\beta_2 X_i}$$

où F^{-1} = inverse de FDC normale standard. (15.9.4)

- Lorsque fréquence relative $\hat{P}_i = 0,66 \rightarrow$ indice d'utilité $I_i = 0,40$.

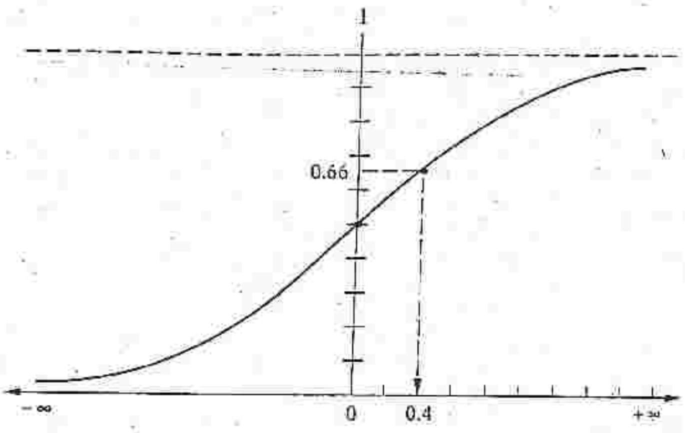


FIGURE 15.5 Normal CDF.

Ensuite, relativmt aisé d'estimer valeurs de β_1 et β_2.

- Remarque :

  Indice d'utilité inobservable $I_i \equiv$ „ déviation équivalente à la normale " (d.e.n) (ou „ normal equivalent deviation") ou „ normit"

  Comme d.e.n. (ou I_i) < 0 lorsque $P_i < 0,5$, en pratique on rajoute chiffre 5 à la d.e.n $\equiv$ „ probit".

• Influence du revenu sur probabilité de posséder une maison ?

On régresse d.e.n (Ii) et probits (d.e.n. + 5) sur intercepte et revenu.

TABLE 15.11
Dependent Variable: I

Variable	Coefficient	Std. error	t statistic	Probability
C	-1.0166	0.0572	-17.7473	1.0397E-07
Income	0.04846	0.00247	19.5585	4.8547E-08

$R^2 = 0.97951$ Durbin-Watson statistic = 0.91384

TABLE 15.12
Dependent Variable: Probit

Variable	Coefficient	Std. error	t statistic	Probability
C	3.9833	0.05728	69.5336	2.03737E-12
Income	0.04846	0.00247	19.5585	4.8547E-08

$R^2 = 0.9795$ Durbin-Watson statistic = 0.9138

Hormis valeur des interceptes, résultats des 2 régressions sont équivalents.

Coeff. de pente = 0,048 → lorsque revenu ↑ d'1 unité, indice d'utilité (Ii) ↑ de 0,048.

Intercepte de la 2ème régr. (3,9833) = valeur intercepte 1ère régr. (-1,0166) + 5. Pas étonnant car intercepte mesure valeur moyenne de la variable dpdte lorsque le revenu est nul.

• Quid de l'impact d'un chgmt unitaire dans le revenu (en milliers de \$) sur la prob. que $Y = 1$ (càd. qu'une famille possède une maison) ?

Nous savons que :

$$\boxed{P_i = P(Y_i = 1 / X_i) = F(\beta_1 + \beta_2 X_i)} \qquad (15.9.2)$$

Il faut dériver cette expression // X_i pour connaître le tx de variation de P_i // au revenu.

En utilisant la „chain rule":

$$\frac{dP_i}{dX_i} = \frac{dF(t)}{dt} \cdot \frac{dt}{dX}$$

où $t = \beta_1 + \beta_2 X_i$

Comme $dF(t)/dt = f(t)$:

$$\boxed{\frac{dP_i}{dX_i} = f(\beta_1 + \beta_2 X_i) \cdot \beta_2}$$

où $f(\beta_1 + \beta_2 X_i)$ est la fct° de densité de probabilité standard normale évaluée en $\beta_1 + \beta_2 X_i$.

Taux de variation de la prob. P_i par rapport au revenu X_i dépend de la valeur particulière prise par X_i.

- Rappel des résultats :

$$\hat{I}_i = -1{,}0166 + 0{,}04846 \; X_i$$
$$\quad (-17{,}7473) \qquad (19{,}5585) \qquad \leftarrow \text{t-stat}$$

Supposons que $X = 6$ (milliers de \$)

$$\Rightarrow f(\hat{\beta}_1 + \hat{\beta}_2 X_i) = f(-1{,}0166 + 0{,}04846 \,.\, 6)$$
$$= f(-0{,}72548)$$

En se référant à des tables statist., pour $z = -0{,}72548$, on trouve que la densité normale standard $= 0{,}3066$.

$$\frac{dP_i}{dX_i} = f(\hat{\beta}_1 + \hat{\beta}_2 X_i) \,.\, \hat{\beta}_2$$

$$= 0{,}3066 \,.\, 0{,}04846 \qquad (\text{pour } X = 6)$$
$$= 0{,}01485$$

Signification ?

En partant d'un revenu de 6000 \$, lorsque celui-ci augmente de 1000 \$, la prob. qu'une famille possède une maison ↗ de 1,4 %.

Résultat semblable à celui obtenu en estimant un modèle logit avec m̂ données (cf. Tab. 15.6, slide 15.33) (calcul + fastidieux avec probit)

- <u>Quid de la probabilité estimée de possession d'une maison pour ≠ niveaux de revenu</u> ?

Nous savons que :

$$P_i = P(Y_i = 1 / X_i) = F(\beta_1 + \beta_2 X_i) \qquad (15.9.2)$$

$$\Rightarrow \boxed{\hat{P}_i = F(\hat{\beta}_1 + \hat{\beta}_2 X_i)}$$

où F est FDC normale standard

$\hat{I}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ (d.e.n)

. Résultats de notre rég. :

$$\hat{I}_i = -1,0166 + 0,04846\ X_i$$

<u>Tab.</u> Valeurs estimées pour la d.e.n („déviation équivalente à la normale" ou $\hat{I}_i$) pour ≠ de X :

X	6	8	10	13	15	20	30	35	40
$\hat{I}_i$, d.e.n. estimée	−0,72	−0,63	−0,53	−0,39	−0,29	0,19	0,43	0,68	0,92

Rem : si X = 8, d.e.n. = −1,0166 + 0,04846 . 8 = −0,63

A partir des d.e.n. estimées, on peut calculer les prob. cumulées pour ≠ nvx de revenus (car $P_i = F(\text{d.e.n.})$)

<u>Exemple</u> : si X = 8 ⟶ d.e.n. = −0,63 ⟶ $\hat{P}_i = 0,2647$
(car $F(-0,63) = 1 - F(0,63) = 1 - 0,7353 = 0,2647$)
Significat° ? Famille ayant revenu de 8000 \$ possèdent une maison avec prob. de 26,47 %.

Comment procéder pour obtenir l'indice I_i et estimer les paramètres β_1 et β_2 ?

→ Réponse dépend des données (groupées ou individuelles).

A) L'estimation du modèle probit avec des données groupées : gprobit

Familles groupées selon niveau de revenu X_i. Pour chque niveau de revenu X_i, il y a N_i familles dont n_i sont propriét. d'une maison ($n_i \leq N_i$).

TABLE 15.10 ESTIMATING THE INDEX I_i FROM THE STANDARD NORMAL CDF

X_i (milliers \$)	N_i	n_i	$\hat{P}_i$	$I_i = F^{-1}(\hat{P}_i)$
6	40	8	0.20	−0.8416
8	50	12	0.24	−0.7063
10	60	18	0.30	−0.5244
13	80	28	0.35	−0.3853
15	100	45	0.45	−0.1257
20	70	36	0.51	0.0251
25	65	39	0.60	0.2533
30	50	33	0.66	0.4125
35	40	30	0.75	0.6745
40	25	20	0.80	0.8416

Notes: (1) $\hat{P}_i$ are from Table 15.5; (2) I_i are estimated from the standard normal CDF.

Calcul de la fréquence relative (càd mesure empirique de P_i) de possession d'une maison pour ≠ nux de revenus :

$$\boxed{\hat{P}_i = \frac{n_i}{N_i}}$$

Utilisation de $\hat{P}_i$ pour obtenir valeur de I_i :

$$\boxed{I_i = F^{-1}(\hat{P}_i)}$$

→ indice d'utilité I_i correspond à l'inverse de la FDC normale standard étant donné $\hat{P}_i$.

si $X = 30 \longrightarrow$ d.e.n. (estimé) $= 0{,}43 \longrightarrow \hat{P}_i = 0{,}6691$

etc...

Les estimations sont assez proches des proportions réelles de familles possédant une maison par classe de revenu (cf. Tab. 15.5, slide 15.30) $\longrightarrow$ modèle estimé assez bon.

B) Le modèle probit avec des données non groupées ou individuelles

GRADE $\sim$ cst, GPA, TUCE, PSI

Analyse de la réussite d'un étudiant à un examen de micro-éco

TABLE 15.7 DATA ON THE EFFECT OF PERSONALIZED SYSTEM OF INSTRUCTION (PSI) ON COURSE GRADES

Observation	GPA grade	TUCE grade	PSI	Grade	Letter grade	Observation	GPA grade	TUCE grade	PSI	Grade	Letter grade
1	2.66	20	0	0	C	17	2.75	25	0	0	C
2	2.89	22	0	0	B	18	2.83	19	0	0	C
3	3.28	24	0	0	B	19	3.12	23	1	0	B
4	2.92	12	0	0	B	20	3.16	25	1	1	A
5	4.00	21	0	1	A	21	2.06	22	1	0	C
6	2.86	17	0	0	B	22	3.62	28	1	1	A
7	2.76	17	0	0	B	23	2.89	14	1	0	C
8	2.87	21	0	0	B	24	3.51	26	1	0	B
9	3.03	25	0	0	C	25	3.54	24	1	1	A
10	3.92	29	0	1	A	26	2.83	27	1	1	A
11	2.63	20	0	0	C	27	3.39	17	1	1	A
12	3.32	23	0	0	B	28	2.67	24	1	0	B
13	3.57	23	0	0	B	29	3.65	21	1	1	A
14	3.26	25	0	1	A	30	4.00	23	1	1	A
15	3.53	26	0	0	B	31	3.10	21	1	0	C
16	2.74	19	0	0	B	32	2.39	19	1	1	A

Notes: Grade $Y = 1$ if the final grade is A
$= 0$ if the final grade is B or C
TUCE = score on an examination given at the beginning of the term to test entering knowledge of macroeconomics
PSI = 1 if the new teaching method is used
= 0 otherwise
GPA = the entering grade point average
Source: L. Spector and M. Mazzeo, "Probit Analysis and Economic Education," *Journal of Economic Education*, vol. 11, 1980, pp. 37–44.

GRADE, $Y = 1$ si étudiant a obtenu „A" pour exam de micro. (réussite)
$= 0$ si étudiant a obtenu „B" ou „C" pour ex. de micro.

GPA = note moyenne de l'étudiant lors de son examen d'entrée

TUCE = score de l'étudiant à un examen de macro au début du trimestre.

PSI = 1 si nlle méthode d'enseignmt utilisée

Modèle probit correspondant :

$$\boxed{L_i = \ln\left(\frac{P_i}{1-P_i}\right) = \beta_1 + \beta_2\, GPA_i + \beta_3\, TUCE_i + \beta_4\, SPI_i + u_i}$$

où $P_i = 1$ si étudiant a réussi son examen (GRADE = 1)

$= 0$ sinon (GRADE = 0)

Méthode d'estimation : (comme pour modèle logit avec données individuelles) procédure non linéaire basée sur la méthode du maximum de vraisemblance.

TABLE 15.13 Dependent Variable: grade
Method: ML-binary probit
Convergence achieved after 5 iterations

Variable	Coefficient	Std. error	Z statistic	Probability
C	-7.4523	2.5424	-2.9311	0.0033
GPA	1.6258	0.6938	2.3430	0.0191
TUCE	0.0517	0.0838	0.6166	0.5374
PSI	1.4263	5950	2.3970	0.0165

LR statistic (3 df) = 15.5458 McFadden R^2 = 0.3774
Probability (LR stat) = 0.0014

Qualitativmt résultats du probit comparables à ceux du logit (cf. Tab. 15.8, slide 15.36)

→ variables GPA et PSI à nouveau signif. à l'inverse de la variable TUCE.

→ ratio de vraisemblance = 15,5458 (p-value = 0,0014) on peut rejeter hyp. que ts les coeff. de pente simult. nuls.

iii) <u>Dans modèle logit</u> : $\boxed{\beta_2 = \dfrac{\Delta\left[\ln\left(\frac{P_i}{1-P_i}\right)\right]}{\Delta X_i}}$

Coeff. de pente mesure variation du log des chances suite à un chgmt unitaire de la valeur d'une variable explicative, ttes les autres variables demeurant constantes.

Taux de variat° de la probabilité d'occurence d'un événement :

$$\boxed{\frac{dP_i}{dX_j} = \beta_j \, P_i\,(1-P_i)}$$

où $\begin{cases} \beta_j = \text{coeff. de rég. partielle du j-ème régresseur} \\ P_i = \dfrac{1}{1+e^{-(\beta_1+\beta_2 X_{2i}+\dots+\beta_k X_{ki})}} \end{cases}$

$\Rightarrow$ Tx de variation de la prob. // X_j dépend non slmt de β_j mais aussi du niveau de prob P_i à partir duquel la variation est mesurée.

iv) <u>Dans modèle probit</u> : $\boxed{\beta_2 = \dfrac{\Delta\left[\ln\left(\frac{P_i}{1-P_i}\right)\right]}{\Delta X_i}}$

Coeff. de pente mesure variation du log des chances suite à un chgmt unitaire de la valeur d'une variable explicative, ttes les 8 variables demeurant constantes.

Taux de variat° de la prob. d'occurence d'un événement assez complexe à calculer.

$$\boxed{\frac{dP_i}{dX_j} = \beta_j \cdot f(Z_i)}$$

où $\begin{cases} f(Z_i) \text{ est la fct° de densité de la variable standard normale} \\ Z_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} \\ (Z_i \text{ correspond au modèle de rég. utilisé ds l'analyse}) \end{cases}$

$\Rightarrow$ Taux de variation de la prob. par rapport à X_j dépend non seulmt de β_j mais aussi des valeurs particulières prises par variables X_2 et X_3

<u>En résumé</u>, ds les modèles logit et probit, tous les régresseurs sont impliqués dans le calcul du tx de variat° de la prob. d'occurence d'un événement.

En revanche, dans le MPL, seul le coeff. de pente associé au régresseur qui change est impliqué dans le calcul du tx de variat° de la prob.

Cette ≠ explique la popularité précoce du MPL. Aujourd'hui facilité de calcul du tx de variat° de la prob. n'est plus un argument en faveur du MPL car effets marginaux des modèles probit et logit sont fournis automat. par plupart de logiciels économ.

Par défaut, +part de logiciels écon. calculent effets marginaux aux valeurs moyennes des var. explicatives $\rightarrow$ ds cas du :

* logit : $d\hat{P}_i / dX_j = \hat{\beta}_j \, \hat{P}_i (1 - \hat{P}_i)$ avec $\hat{P}_i = \dfrac{1}{1 + e^{-(\hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \dots + \hat{\beta}_k \bar{X}_k)}}$
* probit : $d\hat{P}_i / dX_j = \hat{\beta}_j \, f(\hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \dots + \hat{\beta}_k \bar{X}_k)$ où $f(.)$ est fct° de densité normale stand

C) L'effet marginal de la variation d'une unité dans la valeur d'un régresseur dans les divers modèles de régression

Rappel concernant interprétation du coeff. de pente dans 4 modèles suivants :

i) Le modèle de régression linéaire.

ii) Le MPL.

iii) Le modèle logit.

iv) Le modèle probit.

i) Dans modèle de régression linéaire : $\boxed{\beta_2 = \frac{\Delta Y}{\Delta X}}$

Coeff. de pente mesure directmt la variat° de la valeur moyenne de la variable dépdte suite à un chgmt unitaire de la valeur de la variable explicative, ttes les autres variables demeurant constantes.

ii) Dans modèle de probabilité linéaire (MPL) : $\boxed{\beta_2 = \frac{\Delta P}{\Delta X}}$

Coeff. de pente mesure directmt la variat° de la probabilité d'occurence d'un événement suite à un chgmt unitaire de la valeur d'une variable explicative, ttes les autres variables demeurant constantes.

3.3.10. Les modèles logit et probit

Entre les modèles logit et probit, lequel faut-il choisir ?
Dans +part des applications, modèles logit et probit fournissent des résultats semblables.

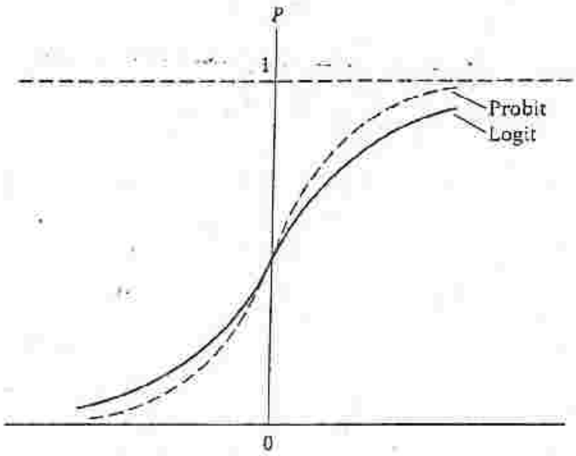


FIGURE 15.6 Logit and probit cumulative distributions.

Mais, il se différencient car distribut° logistique possède des queues de distribution plus grosses → prob. cond. ds modèle logit converge moins vite vers 0 et 1 que ds modèle probit.

TABLE 15.15 VALUES OF CUMULATIVE PROBABILITY FUNCTIONS

	Cumulative normal	Cumulative logistic
Z	$P_1(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-s^2/2} ds$	$P_2(Z) = \frac{1}{1+e^{-z}}$
−3.0	0.0013	0.0474
−2.0	0.0228	0.1192
−1.5	0.0668	0.1824
−1.0	0.1587	0.2689
−0.5	0.3085	0.3775
0	0.5000	0.5000
0.5	0.6915	0.6225
1.0	0.8413	0.7311
1.5	0.9332	0.8176
2.0	0.9772	0.8808
3.0	0.9987	0.9526

Bien que les 2 modèles soient similaires, il faut être prudent ds l'interprétation des coeff. estimés par les 2 modèles.

Dans exemple // réussite d'un examen de microéco. par des étudiants, coeff. associé à la variable GPA (note moyenne lors de l'exam. d'entrée) égal à :

→ 1,6258 ds modèle probit.
→ 2,8261 ds modèle logit.

Raison de cette différence ?

Bien que les dist. log. (à la base du logit) et normales standard (à la base du probit) aient la m̂ moyenne nulle, leurs variances sont différentes ($\sigma^2 = 1$ pour dist. normale standard et $\sigma^2 = \pi^2/3$ pour dist. logistique, où $\pi \simeq 22/7$).

⇒ coeff. du probit × 1,81 ≃ coeff. du logit

où $1{,}81 \simeq \pi/\sqrt{3}$

Dans notre exemple :

Coeff associé à GPA ds probit = 1,6258
→ 1,6258 . 1,81 = 2,94 (≃ coeff ds logit).
Coeff associé à GPA ds logit = 2,8261
→ 2,8261 . 0,55 = 1,55 (≃ coeff ds probit).
(0,55 = 1/1,81)

Cependant, d'après Amemiya (1981) :

$$\boxed{\begin{aligned} \beta_{probit} &= 0{,}625\, \beta_{logit} \\ \beta_{logit} &= 1{,}6\, \beta_{probit} \end{aligned}}$$

$(1{,}6 = 1/0{,}625)$

et

$$\boxed{\begin{aligned} \beta_{MPL} &= 0{,}25\, \beta_{logit} && \text{pour coeff. de pente} \\ \beta_{MPL} &= 0{,}25\, \beta_{logit} + 0{,}5 && \text{pour l'intercepte.} \end{aligned}}$$

3.3.11. Le modèle Tobit

Modèle Tobit = prolongement du modèle probit qui a été développé par James Tobin.

Illustration : exemple concernant possession d'une maison

Dans modèle probit, souci = estimer prob. de posséder une maison en relation avec d'ô variables socio-éco.

Dans modèle tobit, souci = trouver montant financier qu'un individu ou une famille dépense à l'achat d'une maison en fct° de variables socio-éco.

↳ dilemme : on dispose d'information sur les dépenses en logement que pour individus qui achètent réellmt une maison.

- Consommateurs divisés en 2 groupes :
  1) n_1 individus pour lesquels nous disposons d'info sur les régresseurs (ex : revenu, tx d'intérêt du crédit immob., composition de la famille, etc.) et sur la variable dépendante (le montant de la dépense pour le logement).

  2) n_2 individus sur lequels uniq. info sur les régresseurs mais pas sur la var. dpdte.

- Echantillon „censuré" = éch. pour lequel info sur la variable dpdte n'est disponible que pour certaines observations.

  → „modèle probit" ou „modèle de régression censuré" ou „modèle de régr. à variable dpdte limitée" (en raison de la restrict° imposée sur valeurs prises par var. dpdte).

  ≠ Echantillon „tronqué" = éch. ds lequel info sur les régresseurs n'est observée qui si var. dpdte est observée.

- Statistiqmt, modèle s'écrit comme suit :

$$\boxed{\begin{aligned} Y_i &= \beta_1 + \beta_2 X_i + u_i && \text{si MD} > 0 \\ &= 0 && \text{autrement} \end{aligned}} \qquad (15.11.1)$$

  où MD = mbre de droite de l'égalité

- Comment procéder pour estimer modèle tobit ?

> Meilleure procédure : utilisation de la méthode du max de vraisemblance. Assez complexe mais nbrx logiciels écon. permettent de l'appliquer.

> Alternative plus simple : procédure d'estimation en 2 étapes proposée par James Heckman.

Première étape : estimer prob. qu'un cons. achète une maison à l'aide d'un modèle probit.

Deuxième étape : estimer modèle de départ ($Y_i = \beta_1 + \beta_2 X_i$) en ajoutant une variable (le „ratio inverse de Mills" ou le „ tx de hasard" qui est dérivé du probit)

Procédure de Heckman fournit des estimateurs consistants des paramètres inconnus, mais pas aussi efficients que ceux du maximum de vraisemblance.

- Peut-on estimer ce modèle de régression (15.11.1) en utilisant uniq. n_1 observations (pour lesquelles info sur régresseurs et var dpdte dispo) sans se soucier des n_2 restantes (pour lequelles uniq info sur régresseurs pas sur var dpdte) ?

NON !

Pq ?

Estimateurs par MCO des paramètres obtenus du sous-ensemble n_1 seront biaisés ($\hat{m}$ asympt.) et inconsistants.

3.3.12. Compléments d'analyse sur les modèles de régression à variable qualitative

Trois autres types de modèles :

i) les modèles logit et probit ordinaux

ii) Les modèles logit et probit multinomiaux

iii) les modèles de durée

A) Les modèles logit et probit ordinaux

Dans modèles logit et probit bi-variés, on cherche à modéliser une variable de réponse (càd. qualitative dépdte) par oui ou non.

Pblme : svt variable de réponse présente + de 2 résultats ordinaux par nature → résultats peuvent être ordonnés (classés) mais ils ne pvt être représentés sur une échelle d'intervalles.

Exemple : enquête où réponses du type „tout à fait d'accord", „moyennement d'accord" ou „pas du tout d'accord". Variable de réponse ordinale car ordre logique ds les réponses mais elles ne pvt être représentées sur échelle d'intervalles.

⇒ modèles logit et probit ordinaux

B) Les modèles logit et probit multinomiaux

Dans certains cas, la variable de réponse n'est pas ordonnée.

Exemples: • choix du mode de transport pour se rendre au travail (bicyclette, moto, voiture, métro, ...)
→ pas d'ordre clair (logique) dans les réponses
→ réponses nominales.

• affiliation sectorielle des travailleurs (primaire, secondaire, tertiaire)
→ variable dpdte nominale.

⇒ modèles probit et logit multinomiaux.

C) Les modèles de durée

Quid lorsque questions du type:

i) Quels sont les déterminants des périodes de chômage?
ii) Quelles sont les raisons de la durée de vie d'une ampoule électrique?
iii) Quels facteurs déterminent la durée d'une grève?

⇒ modèles de durée où variable dépdte = variable aléatoire mesurant durée d'un événement.

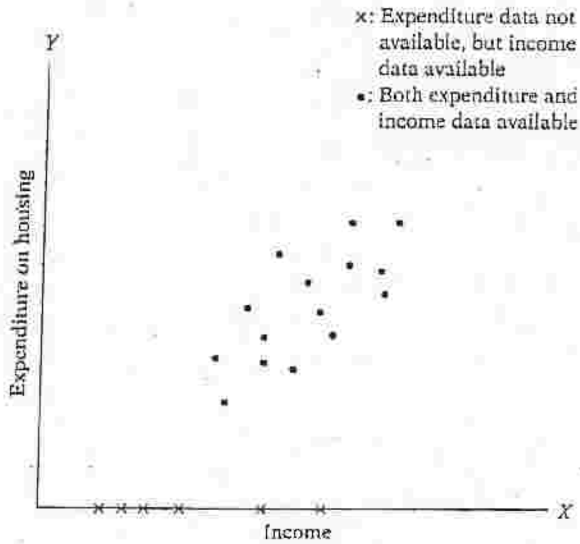


FIGURE 15.7 Plot of amount of money consumer spends in buying a house versus income.

- Ttes les obs. (n_2) pour lesquelles Y n'est pas observé sont notées par des croix sur l'axe horizontale.

  Ttes les obs. (n_2) pour lesquelles Y est observé sont notées par des points et se répartissent ds plan X-Y.

, Evident que si on estime droite de régression uniq. avec n_1 obs., intercepte et coeff. de pente seront ≠ de ceux impliquant ttes les obs. ($n_1 + n_2$).