

Économétrie II

L3 Économétrie – L3 MASS

Ch. 7. Variables Dépendantes Dichotomiques

Prof. Philippe Polomé, U. Lyon 2

Année 2015-2016

Table des matières

Ch. 7. Variables Dépendantes Dichotomiques

Définition, interprétation & MCO

MCO avec VDL dichotomique

Modèles à VDD : logit & probit

Interprétation des modèles Logit & Probit & exemple

Autres types de VDL et conclusions

Rappel

1. ✓ $E(\varepsilon_t) = 0 \forall t$: **Espérance nulle**
2. ✓ $var(\varepsilon_t) = \sigma^2 \forall t$: **Homoscédasticité**
3. ✓ $cov(\varepsilon_t, \varepsilon_s) = 0 \forall t \neq s$: **Pas d'autocorrélation**
4. ✓ $E(\varepsilon_t x_t) = 0 \forall t$: **Exogénéité**
5. ✓ La matrice X est de plein rang : **Pas de multicolinéarité**
6. ✓ Le modèle est **correctement spécifié**
7. **La variable dépendante Y est continue**

Table des matières

Ch. 7. Variables Dépendantes Dichotomiques

Définition, interprétation & MCO

MCO avec VDL dichotomique

Modèles à VDD : logit & probit

Interprétation des modèles Logit & Probit & exemple

Autres types de VDL et conclusions

Variable dépendante limitée VDL

- ▶ **Binaire/dichotomique** : 2 valeurs seulement
- ▶ **Catégorique** : nombre fini de valeurs
 - ▶ p.e. codée de 1 à 3 “faible”, “moyen”, “élevé” (ordinaire)
 - ▶ “bus”, “voiture”, “vélo” (non-ordonné)
- ▶ Variable **bornée** par une valeur critique
 - ▶ D'en dessous : donations annuelles par ménages limitées par zéro
 - ▶ Par le haut : nombre de spectateurs limité par la capacité du stade
 - ▶ Des 2 côtés : Notes de cours entre 0 et 20 en France

Interprétation

- ▶ Une variable continue a un sens quantitatif (cardinal)
- ▶ Une VDL a souvent un sens qualitatif
 - ▶ p.e. satisfaction 5="great", 4="good", 3="fair", 2="poor", 1="very poor".
 - ▶ Satisfaction = 5 est mieux que 4, qui est mieux que 3
 - ▶ Mais satisfaction = 5 n'est pas 5 fois plus que satisfaction = 1
 - ▶ La différence entre satisfaction = 4 et satisfaction = 3 n'est pas la même qu'entre 3 et 2

MCO

- ▶ Si on utilise MCO pour analyser l'impact de régresseurs sur une VDL, on la traite comme quantitative
- ▶ Comme une relation linéaire entre la VDL est assumée, MCO impose des effets marginaux CONSTANTS quel que soit le niveau du régresseur → 2 défauts :
 1. MCO peut prédire des valeurs de VDL + petites que leur min possible ou + grdes que leur max possible
 2. MCO peut prédire des effets marginaux + grds que le + grd changements possibles qui peut affecter la VDL, p.e. 1.2 pour une VDL dichotomique
- ▶ EMV – **Estimateur du Maximum de Vraisemblance** (MLE) évite ces défauts en permettant une relation **non-linéaire** entre la VDL et ses régresseurs

Table des matières

Ch. 7. Variables Dépendantes Dichotomiques

Définition, interprétation & MCO

MCO avec VDL dichotomique

Modèles à VDD : logit & probit

Interprétation des modèles Logit & Probit & exemple

Autres types de VDL et conclusions

Modèle à proba linéaire

- ▶ Soit y VDL dichotomique (VDD)
- ▶ Que signifie un MRL dans ce cas ?

$$y = X\beta + \varepsilon$$

- ▶ β_j ne peut être interprété comme le changement *ceteris paribus* en y étant donné un changement unitaire en x_j puisque y ne peut changer que de zéro à un ou de un à zéro

Interprétation de β

- ▶ Sous hypothèse d'exogénéité $E(\varepsilon|X) = 0$, on a :

$$E(y|X) = \Pr\{y = 1|X\} = X\beta$$

- ▶ La proba de succès $\Pr\{y = 1|X\}$ ou “proba de réponse”, est une fonction linéaire de x_j
- ▶ Le MRL avec VDD est appelé **Modèle à Proba Linéaire (MPL)**
- ▶ Que mesure β_j dans le MPL ?
 - ▶ Le changement ceteris paribus en la proba de succès quand x_j change
 - ▶ $\Delta P(y = 1|\Delta x_j) = \beta_j \Delta x_j$
- ▶ Si on écrit la valeur ajustée \hat{y} (la proba de succès prédite) comme $\hat{y} = X\hat{\beta}$
 - ▶ $\hat{\beta}_j$ mesure le changement prédit en la proba de succès lorsque x_j augmente d'une unité

Exemple : Participation des femmes à l'offre de travail

- ▶ Échantillon aléatoire de 753 femmes
 - ▶ Wooldridge : Gretl [mroz](#)
- ▶ Estime par MCO la proba qu'une femme travaille contre salaire
 - ▶ ou soit en recherche d'emploi en principe
- ▶ La variable dépendante est dichotomique
 - ▶ 1 si la femme travaille contre salaire hors de chez elle à un point quelconque de l'année
 - ▶ 0 autrement
- ▶ Estimation par MCO
 - ▶ pour les 753 femmes de l'échantillon, 16 ont des valeurs ajustées < 0 et 17 ont > 1

Impact des jeunes enfants

- ▶ Une des variables explicatives est “nbr d’enfants de moins de 6 ans”, kid<6
- ▶ Après estimation par MCO
 - ▶ le coefficient de “kid<6” est significatif et egal à -0.262
- ▶ Donc : un enfant de moins de 6 ans en plus réduit
 - ▶ la proba de participation de -0.262 - ceteris paribus (sous exogénéité)
 - ▶ **quel que soit** le nombre d’enfants de moins de 6 ans
 - ▶ Passer de 0 à 4 enfants de moins de 6 ans réduit la proba de travailler de $0.262 * 4 = 1.048$ (impossible)

Table des matières

Ch. 7. Variables Dépendantes Dichotomiques

Définition, interprétation & MCO

MCO avec VDL dichotomique

Modèles à VDD : logit & probit

Interprétation des modèles Logit & Probit & exemple

Autres types de VDL et conclusions

Spécification

- ▶ Soit une classe de modèles non-linéaires à réponse dichotomique :

$$\Pr(y = 1|X) = G(X\beta)$$

G est une fonction qui prend des valeurs strictement entre zéro et un : $0 \leq G(z) \leq 1, \forall$ nombre réel z

- ▶ Ceci garanti que les proba estimées de réponse seront strictment entre zéro et un
- ▶ Implique une relation **non linéaire** entre la variable dépendante et les régresseurs
- ▶ De nombreuses fonctions sont candidates
 - ▶ 2 sont populaires : logistique et normale

Logit et Probit

- ▶ Modèle logit, G est la fonction de distribution (densité cumulative) d'une v.a. logistique standard :

$$G(z) = \exp(z) / [1 + \exp(z)] = \Lambda(z)$$

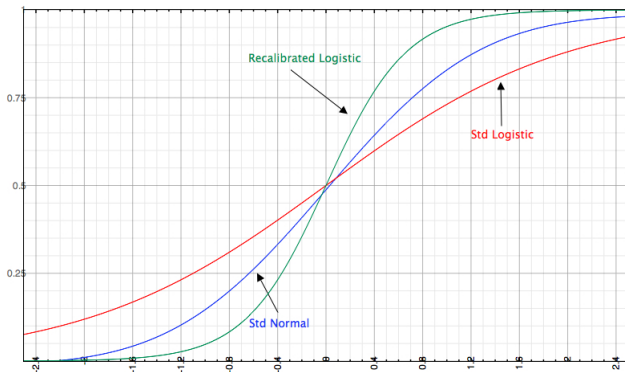
- ▶ Modèle probit, G est la fonction de distribution d'une v.a. normale standard, dont on note la densité $\phi(\cdot)$:

$$G(z) = \int_{-\infty}^z \phi(t) dt$$

avec $\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2)$

Logit vs. Probit

- ▶ Les distributions logistique et normale sont similaires
- ▶ La logistique rend les calculs plus simples et permet des simplifications importantes dans des modèles plus avancés



Modèle à Variable Latente

- ▶ Soit y^* une variable latente (c'est-à-dire pas observée directement) t.q.

$$y^* = X\beta + \varepsilon$$

- ▶ P.e. y^* est l'utilité d'acheter une nouvelle voiture
- ▶ Logit et probit peuvent être obtenus d'un modèle à variable latente qui satisfait toutes les hypothèses du MRL classique
- ▶ On n'observe pas l'utilité, mais seulement la conséquence de la décision individuelle

$$\begin{cases} y_i^* < 0 \implies y_i = 0 \\ y_i^* \geq 0 \implies y_i = 1 \end{cases}$$

- ▶ On observe que la personne a ($y = 1$) ou n'a pas ($y = 0$) acheté de nvlle voiture

Proba de réponse

- ▶ Hypothèse : ε est indépendant de x et ε est soit standard logistique soit standard normale
- ▶ On dérive les proba de réponse pour y :

$$\begin{aligned}\Pr\{y = 1|X\} &= \Pr\{y^* \geq 0|X\} \\ &= \Pr\{\varepsilon > -(X\beta)|X\} \\ &= 1 - G(-(X\beta)) \\ &= G(X\beta)\end{aligned}$$

- ▶ Comme ε est normale ou logistique, elle est sym autour zéro, donc $1 - G(-z) = G(z) \forall$ nbr réel z

Estimation Maximum de Vraisemblance

- ▶ Estimation des modèles Logit et Probit par Maximum de Vraisemblance (Maximum Likelihood)
- ▶ Dans ce cours, MV = boîte noire qui produit des estimations $\hat{\beta}_{MV}$
 - ▶ Consistantes mais biaisées
 - ▶ Asymptotiquement efficaces
 - ▶ Asymptotiquement normales
 - ▶ Pour autant que les hypothèses du modèles soient vraies
 - ▶ Sinon, c'est pas forcément fichu, mais c'est compliqué
 - ▶ En pratique : on donne au logiciel y comme variable dépendante et X (plein rang) comme régresseur
- ▶ MV sera vu en détail en M1

Table des matières

Ch. 7. Variables Dépendantes Dichotomiques

Définition, interprétation & MCO

MCO avec VDL dichotomique

Modèles à VDD : logit & probit

Interprétation des modèles Logit & Probit & exemple

Autres types de VDL et conclusions

Cas d'un régresseur continu x_j

- ▶ L'effet d'un changement marginal en x_j sur la proba de réponse $\Pr\{y = 1|X\} = p(X)$ est donné par la dérivée partielle

$$\frac{\partial p(X)}{\partial x_j} = \frac{\partial G(X\beta)}{\partial x_j} = g(X\beta)\beta_j$$

- ▶ C'est l'**effet marginal** de x_j : il dépend des valeurs prises par **tous** les régresseurs (pas seulement de x_j)
 - ▶ On peut le calculer en des points “intéressants” de X , p.e. moyennes sur l'échantillon
 - ▶ Qd un régresseur est discret sa moyenne sur l'échantillon ne veut pas dire grd chose
 - ▶ Ou bien on peut le calculer pour chaque i dans l'échantillon et calculer la moyenne des effets marginaux “individuels”
 - ▶ En général ça ne coïncide pas

Effets marginaux Logit – Probit

- ▶ Dans Logit $g(z) = \frac{\exp(z)}{[1 + \exp(z)]^2}$ et $g(0) = .25$
- ▶ Dans Probit $g(z) = \phi(z)$ (la densité normale standard) et $\phi(0) = 1/\sqrt{2\pi} \simeq .4$
- ▶ Ces modèles ne peuvent pas avoir des effets marginaux $g(X\beta) x_j$ plus grand que un

Cas d'un régresseur discret

- ▶ Effet d'un changement en x_j discret
 - ▶ de a à b (souvent, de 0 à 1)
 - ▶ sur la proba de réponse $\Pr\{y = 1|X\} = p(X)$
 - ▶ On écrit X_{-j} l'ensemble des régresseurs sauf x_j , de façon similaire β_{-j}

$$\Delta \hat{p}(X_i) = G\left(X_{-ji}\hat{\beta}_{-j} + b\hat{\beta}_j\right) - G\left(X_{-ji}\hat{\beta}_{-j} + a\hat{\beta}_j\right)$$

- ▶ Un tel effet est différent d'individu à individu
- ▶ Dans Stata
 - ▶ Commande `mfx` après `logit` ou `probit` : effets marginaux de régresseurs continus
 - ▶ Commande `prvalue` pour un régresseur discret

Mesures de qualité d'ajustement

- ▶ Le **pourcentage correctement prédit**
- ▶ $\forall i$ calculer la proba ajustée que y_i prenne la valeur 1, $G(X_i\hat{\beta})$
 - ▶ Si $\geq .5$ on “prédit” $y_i = 1$ et zéro sinon
 - ▶ On calcule le % de prédictions correctes
- ▶ Problème : possible d’obtenir de hauts % correctement prédits sans que le modèle soit bien utile
 - ▶ P.e., ds un échantillon de 200, 180 observations ont $y_i = 0$ dont 150 sont prédites zéro et 20 obs ont $y_i = 1$ toutes prédites zéro
 - ▶ Le modèle est clairement mauvais
 - ▶ Mais on a qd même 75% de prédictions correctes
- ▶ Pour cela, on rapporte un tableau de prédiction 2×2 (p.e. Stata)

Pseudo R-carré

- ▶ $Pseudo - R^2 = 1 - \ln \mathcal{L}_{UR} / \ln \mathcal{L}_0$
 - ▶ $\ln \mathcal{L}_{UR}$ la log-vraisemblance du modèle estimé (voir M1)
 - ▶ $\ln \mathcal{L}_0$ celle du modèle avec seulement l'intercept
- ▶ semblable à R^2 pour la régression MCO
 - ▶ car $R^2 = 1 - SSR_{UR} / SSR_0$
- ▶ Il existe d'autres mesures de qualité de l'ajustement,
 - ▶ mais l'ajustement n'est généralement pas aussi important que la significativité statistique et économique des régresseurs

Exemple : Participation des femmes à l'offre de travail

inlf	=1 if in labor force, 1975
inc*	(faminc - wage*hours)/1000
educ	years of schooling
exp	actual labor mkt exper
exp ²	
age	woman's age in yrs
kid < 6	# kids < 6 years
kid ≥ 6	

Fichier Gretl [MROZ](#) (tab Wooldridge)

Échelle

- ▶ Les **tailles** des coefficients **ne sont pas** directement comparables entre modèles
- ▶ C'est pcq avec les variables dichotomiques y on pourrait multiplier l'ensemble des coefficients par n'importe quelle constante positive sans changer le modèle
 - ▶ \implies la variance de y n'est pas identifiée en général
- ▶ Approximativement,
 - ▶ diviser les estimations logit par 4 et les probit par 2.5 pour les rendre comparables aux estimations du MPL

Modèles pour *inlf* (Participation des femmes à l'offre de travail)

x_j	$\hat{\beta}_{OLS}$	l _{tl}	$\hat{\beta}_{logit}$	l _{tl}	$\hat{\beta}_{probit}$	l _{tl}
inc*	-.003	2.4	-.021	2.5	-.012	2.5
educ	.038	5.2	.221	5.1	.131	5.2
exp	.039	7	.206	6.4	.123	6.7
exp ²	-.000	3.2	-.003	-3.1	-.002	3.2
age	-.016	6.5	-.088	-6.0	-.053	6.2
kid < 6	-.262	7.8	-1.44	-7.1	-.868	7.3
kid ≥ 6	.013	1	.060	.8	.036	.8
Cst	.586	3.8	.425	.5	.27	.5
ln \mathcal{L}	–		-402		-401	
p- R^2	.26	R_a^2	.220		.221	

Comparer les effets marginaux

- ▶ MPL, effets marginaux constants = coef estimés
- ▶ Probit : $\phi \left(\bar{X} \hat{\beta}_{probit} \right) \simeq .4$
 - ▶ \bar{X} qui comprend les moyennes de exp^2 : $\overline{exp^2} \neq e\bar{x}p^2$
 - ▶ Coefficient $\times 0.4$ = effet marginal de la variable sur la proba à la moyenne de l'échantillon
 - ▶ Sans doute différent en d'autres points
 - ▶ Sera différent entre i

Exemple : Participation des femmes à l'offre de travail

- ▶ Une femme avec $inc^* = 20.13$, $educ = 0.13$, $exp = 10.6$, $age = 42.5$
 - ▶ Environ les moyennes de l'échantillon
- ▶ Proba pour $kid < 6 = 0$
 - ▶ Calculer la distribution normale standard $\Phi(\bar{X}\hat{\beta})$ en $kid < 6 = 0$:
Pr = 0.707
- ▶ Baisse estimée de la proba de travailler en passant de zéro à un jeune enfant ?
 - ▶ Calculer la distribution normale standard $\Phi(\bar{X}\hat{\beta})$ en $kid < 6 = 1$:
Pr = 0.373
 - ▶ On a environ $0.373 - 0.707 = -0.334$: proba environ 0.334 plus petite lorsque la femme a un petit enfant
- ▶ Si cette femme passe de 1 à 2 petits enfants, la proba tombe plus bas,
 - ▶ mais l'effet marginal est moindre : $0.117 - 0.373 = -0.256$

Table prédictive (estat classification)

Predict	Model	$y_i = 1$	$y_i = 0$	Total
$\hat{y}_i = 1$	LPM	350	122	472
	Logit	347	118	465
	Probit	348	120	468
$\hat{y}_i = 0$	LPM	78	203	281
	Logit	81	207	288
	Probit	80	205	285
Total		428	325	753

Remarque : Endogénéité

- ▶ Les enfants n'arrivent pas par hasard
- ▶ La décision d'avoir un enfant de plus est prise en même temps que celle de participer au marché du travail
 - ▶ Voire, un changement professionnel peut mener à réajuster la vie familiale
- ▶ Donc : le nombre d'enfants est endogène dans cette régression
 - ▶ Sans doute, particulièrement le nombre de jeunes enfants
- ▶ Deux équations latentes
 - ▶ Utilités
$$\begin{aligned} U_1^* &= X_1\beta_1 + \varepsilon_1 \\ U_2^* &= X_2\beta_2 + \varepsilon_2 \end{aligned}$$
 - ▶ On observe
$$Y_1 = \begin{cases} 1 & \text{si } U_1^* > 0 \\ 0 & \text{sinon} \end{cases}$$
 - ▶ idem pour U_2^*
 - ▶ Vraisemblablement, ε_1 et ε_2 corrélés

Table des matières

Ch. 7. Variables Dépendantes Dichotomiques

Définition, interprétation & MCO

MCO avec VDL dichotomique

Modèles à VDD : logit & probit

Interprétation des modèles Logit & Probit & exemple

Autres types de VDL et conclusions

3 catégories principales

- ▶ Variables ordinales
- ▶ Variables catégoriques
- ▶ Variables continues bornées

Variables ordinales

- ▶ P.e. “satisfaction” codée 1 à 3
- ▶ On part d’un principe de variables observées y_i et latente y_i^* t.q.

$$\left\{ \begin{array}{lll} y_i = 1 & \text{if} & y^* \leq \kappa_1 \\ y_i = 2 & \text{if} & \kappa_1 \leq y^* \leq \kappa_2 \\ y_i = 3 & \text{if} & \kappa_2 \leq y^* \end{array} \right.$$

où κ_1 et κ_2 sont appelés “**seuils**” ou “points de coupure”

- ▶ Cas des VDL “**ordonnée**” : y a un ordre naturel
- ▶ Estimation par des extensions de probit et logit (dits ordonnés) dans Stata par MV
- ▶ Interpretation des coefficients assez différente
 - ▶ Même interprétation du signe et de la significativité

Variables catégoriques

- ▶ Même principe latent, sauf qu'il n'y a pas d'ordre
 - ▶ On parle de choix multinomial entre **alternatives** A, B, C...
- ▶ Logit/probit dits multinomial
 - ▶ À coefficients fixes entre alternatives (logit/probit conditionnel) : seuls les régresseurs qui changent entre les alternatives
 - ▶ À coefficients variables entre alternatives : on peut estimer un coef pour des régresseurs constants entre alternatives
- ▶ Interprétation également assez différente
 - ▶ Même interprétation du signe et de la significativité

Conclusions

1. On préfère MV à MC en présence de VDL
2. Les effets marginaux sont moins évidents à calculer parce que les modèles deviennent non-linéaires
3. Les modèles à VDL sont nombreux et présentent généralement des interprétations assez différentes les uns des autres
 - 3.1 Du point de vue logiciel il ne s'agit jamais que d'une variable dépendante et de régresseurs
 - 3.2 Les interprétations du signe et de la significativité restent les mêmes