



HAL
open science

L'économétrie des variables qualitatives

Gérard Lassibille

► **To cite this version:**

Gérard Lassibille. L'économétrie des variables qualitatives. IREDU, 28, 64 p., 1978, Les Cahiers de l'IREDU, 2-85634-028-8. halshs-02046081


HAL Id: halshs-02046081

<https://shs.hal.science/halshs-02046081v1>

Submitted on 22 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**institut de
recherche sur
l'économie de
l'éducation**

centre national de la
recherche scientifique

gérard lassibille

**L'ECONOMETRIE
DES VARIABLES QUALITATIVES**

université de dijon - faculté de science économique et de gestion

adresse postale: centre universitaire - bâtiment sciences mirande
21000 dijon - tél. (80) 65.44.56

L'ÉCONOMÉTRIE DES VARIABLES
QUALITATIVES

Gérard LASSIBILLE
Attaché de recherche C.N.R.S.

Cahier de l'IREDU n° 28
N° ISBN 2-85634-028-8

Juillet 1978

REMERCIEMENTS

Qu'il nous soit permis d'exprimer ici notre reconnaissance à nos amis et collègues de l'Institut de Recherche sur l'Economie de l'Education, qui ont su mettre à notre disposition les données nécessaires à l'estimation des modèles présentés dans cet ouvrage,

Nous remercions également Paul Delannoy et Christian Michelot, assistants du Laboratoire d'Analyse Numérique de l'U.E.R. M. I. P. C. de l'Université de Dijon, qui ont réalisé les programmes informatiques d'optimisation.

INTRODUCTION

En microéconomie, beaucoup de comportements sont qualitatifs. Il arrive alors fréquemment que le chercheur, intéressé par l'explication ou par le pouvoir explicatif de ces phénomènes, utilise comme outil d'analyse la régression multivariée incluant donc des variables qualitatives. Celles-ci peuvent être de deux sortes, dichotomiques d'une part, polytomiques d'autre part, ces dernières constituant une généralisation des précédentes.

A titre d'illustration, considérons l'évènement

$E = \{\text{possession d'un diplôme par un individu}\}$

Si nous nous intéressons exclusivement à la possession ou non d'un diplôme, nous définissons alors une variable dichotomique que nous codons 1 si l'individu possède un diplôme et 0 sinon. Le choix de ces valeurs, de même que leur affectation à une modalité plutôt qu'à l'autre est totalement arbitraire. Seul importe de tenir compte de l'échelle de valeurs ainsi que de l'utilisation des valeurs retenues lors de l'interprétation du problème.

De telles variables peuvent être selon les cas, endogènes ou exogènes. L'estimation d'un modèle comportant des variables qualitatives exclusivement explicatives ne présente aucune difficulté, si ce n'est dans l'interprétation à donner aux estimateurs des paramètres inconnus de ces variables. Imaginons par exemple un modèle linéaire du type

$$y_i = a + bx_i + \varepsilon_i$$

dans lequel y_i est une variable continue, par exemple le revenu de l'individu i et x_i une variable dichotomique représentant la possession ou non d'un diplôme par ce même individu. Dans le cas où x_i prend la valeur 1 lorsque l'individu a un diplôme et la valeur 0 dans le cas contraire, alors l'estimateur \hat{b} du paramètre inconnu b s'interprète comme l'avantage de salaire dont bénéficie un individu diplômé par rapport à un individu non diplômé.

Dans un modèle à variable dépendante dichotomique, l'interprétation à donner à la variation d'une variable exogène qualitative est équivalente, la principale différence réside dans la procédure d'estimation de cette variation puisque les hypothèses généralement admises, qu'elles soient relatives aux erreurs ou à la forme du modèle ne sont plus respectées.

Le but de cet ouvrage est de faire le point sur les solutions apportées à l'estimation de modèle à variable dépendante qualitative (dichotomique ou polytomique).

Les trois premiers chapitres sont consacrés à l'étude du modèle à variable endogène dichotomique. Le chapitre I examine les difficultés d'estimation du modèle linéaire par les méthodes usuelles de régression et conclut, en raison de la nature particulière de la variable expliquée, à l'abandon de la représentation linéaire. A partir de ceci, le chapitre II propose diverses formulations non linéaires permettant de tenir compte du caractère de la variable endogène. Malheureusement les estimations, a priori séduisantes, de ces formulations soulèvent des difficultés économétriques importantes et principalement celles du groupement des données et de la transformation d'un modèle non linéaire en un modèle linéaire. Le chapitre III expose alors une technique d'estimation de données individuelles sous l'hypothèse d'une

représentation logistique du phénomène, par la méthode du maximum de vraisemblance qui offre l'avantage d'éliminer les inconvénients précités.

Les deux derniers chapitres de cet ouvrage généralisent la représentation logistique précédente et la procédure d'estimation du maximum de vraisemblance à deux modèles économétriques particuliers. Ainsi le chapitre IV expose l'estimation du modèle à variable dépendante polytomique, alors que le chapitre V présente un système d'équation simultanée à variables endogènes qualitatives.

Chaque fois que cela nous a été possible, nous avons illustré les différents chapitres de cet ouvrage d'exemples pratiques d'estimation afin de le rendre accessible au lecteur peu familiarisé avec les présentations économétriques théoriques.

CHAPITRE I

LA FONCTION DE PROBABILITE LINEAIRE

Ce chapitre a pour objet l'étude de l'estimation de la probabilité de réalisation d'un évènement E étant donné un certain nombre de caractéristiques associées à cette éventualité, sous l'hypothèse d'une représentation linéaire du phénomène. Le modèle économétrique sous-jacent ne vérifie plus les hypothèses classiques des moindres carrés ordinaires d'une part, la définition de la probabilité d'autre part. Si nous pouvons facilement pallier le premier inconvénient en conservant la forme du modèle, le second, quant à lui incite à la rejeter brutalement.

I.1. CONDITIONS D'ESTIMATION PAR LES MOINDRES CARRÉS ORDINAIRES

Considérons l'évènement

$E = \{ \text{réussite d'un individu à l'examen de fin d'année d'études} \}$

Notons

$y_i = 1$ si cet évènement se réalise pour l'individu i

$y_i = 0$ sinon.

Supposons que la variable y_i , qui dans le cas présent est une variable dichotomique, soit déterminée par k variables exogènes indépendantes et fixes, x (binaires ou non). L'hypothèse la plus

simple que nous pouvons formuler lorsqu'une relation est supposée exister entre un certain nombre de variables est l'hypothèse de linéarité. C'est-à-dire que nous avons le modèle :

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \epsilon_i \quad i = 1, \dots, n$$

où ϵ_i est un terme d'erreur aléatoire additif.

Ce que nous pouvons encore écrire :

$$y_i = x^i \beta + \epsilon_i \quad i = 1, \dots, n$$

où β est le vecteur d'ordre $(k + 1, 1)$ des paramètres inconnus et x^i le vecteur d'ordre $(1, k + 1)$ des variables explicatives associées à l'individu i .

L'estimation des paramètres inconnus du modèle par la méthode des moindres carrés ordinaires n'est valide que si les hypothèses ci-dessous sont vérifiées :

$$E(\epsilon_i) = 0 \quad \forall_i$$

$$E(\epsilon_i \epsilon_j) = \sigma^2 \text{ pour } i = j \\ = 0 \text{ sinon.}$$

C'est-à-dire si les erreurs aléatoires ont une espérance mathématique nulle et sont homoscédastiques et indépendantes. Cette dernière hypothèse revient à formuler que le comportement de l'individu i est indépendant du comportement de l'individu j et que la variance de l'erreur est identique quel que soit l'individu.

Les estimateurs $\hat{\beta}_k$, obtenus par minimisation de la somme des carrés des écarts verticaux, sont linéaires par rapport à la variable endogène y_i . Sous l'hypothèse de nullité de l'espérance mathématique des erreurs ils sont centrés ou sans biais. Sous l'hypothèse d'indépendance et d'homoscédasticité des erreurs, leur variance est

est égale à :

$$\text{Var}(\hat{\beta}_k) = \frac{\sigma^2}{\sum_{k=1}^n x_k^2}$$

$$\text{avec } \sum_{k=1}^n x_k^2 = \frac{1}{n} \sum_{i=1}^n (x_{ki} - \bar{x}_k)^2$$

D'après le théorème de GAUSS-MARKOV, tout autre estimateur $\hat{\beta}_k^*$ de β_k , non biaisé et linéaire par rapport à la variable endogène y_i , possède une variance supérieure à celle de $\hat{\beta}_k$. Aussi les estimateurs des moindres carrés ordinaires sont-ils à variance minimale. Etant de plus linéaires et centrés ce sont alors des estimateurs BLUE (Best Linear Unbiased Estimator). Outre les hypothèses précédentes sur les erreurs aléatoires ϵ_i , il est généralement admis que celles-ci suivent une loi normale d'espérance mathématique nulle et de variance σ^2 . Soit :

$$f(\epsilon_i) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2} \epsilon_i^2 \right\}$$

De cette hypothèse découle la possibilité de tester la pertinence des variables exogènes supposées déterminer le phénomène expliqué et a fortiori celle de construire des intervalles de confiance pour les paramètres inconnus du modèle.

Si l'hypothèse de normalité des erreurs n'est pas acceptable alors les tests de significativité des variables, du type tests de STUDENT sont nécessairement biaisés. Il en est également de même pour les estimateurs des paramètres inconnus du modèle si l'hypothèse d'indépendance et d'homoscédasticité des erreurs n'est pas vérifiée.

I.2. CONSEQUENCES DU CARACTERE DICHOTOMIQUE
DE Y_i

Rappelons que le modèle à estimer est de la forme

$$y_i = x_i^i \beta + \varepsilon_i \quad i = 1, \dots, n$$

dans lequel y_i prend la valeur 1 si l'évènement expliqué se réalise pour l'individu i , et 0 sinon.

La variable endogène y_i , ne pouvant prendre que deux valeurs, est une variable aléatoire indicatrice dont l'espérance mathématique conditionnelle est égale à :

$$E\{y_i | x_i^i\} = \text{Prob}\{y_i = 0 | x_i^i\} \cdot 0 + \text{Prob}\{y_i = 1 | x_i^i\} \cdot 1 = \text{Prob}\{y_i = 1 | x_i^i\}$$

Sous l'hypothèse de nullité de l'espérance mathématique de l'erreur aléatoire ε_i , nous avons :

$$E\{y_i | x_i^i\} = x_i^i \beta$$

D'où

$$\text{Prob}\{y_i = 1 | x_i^i\} = x_i^i \beta$$

En raison du caractère dichotomique de la variable endogène son espérance mathématique n'est rien d'autre que la probabilité conditionnelle de réalisation de l'évènement expliqué étant donné le vecteur des variables exogènes x_i^i (d'où le nom de fonction de probabilité linéaire donné au modèle ci-dessous). Autrement dit, la valeur calculée du modèle est une estimation de la probabilité conditionnelle de réalisation associée à la valeur 1 de la variable endogène.

Il est encore fréquent de voir des estimations d'un tel modèle par la méthode des moindres carrés ordinaires bien que les hypothèses relatives à cette méthode ne soient plus vérifiées. Pour s'en convaincre, calculons la variance du terme aléatoire ϵ_i . Celui-ci ne peut prendre que deux valeurs, à savoir :

$$\begin{aligned}\epsilon_i &= 1 - x^i\beta && \text{si } y_i = 1 \\ \epsilon_i &= -x^i\beta && \text{si } y_i = 0\end{aligned}$$

Sous l'hypothèse de nullité de l'espérance mathématique de l'erreur aléatoire, nous avons :

$$E(\epsilon_i) = \text{Prob}\{\epsilon_i = -x^i\beta\} (-x^i\beta) + \text{Prob}\{\epsilon_i = 1-x^i\beta\} (1-x^i\beta) = 0$$

Or

$$\text{Prob}\{\epsilon_i = -x^i\beta\} + \text{Prob}\{\epsilon_i = 1-x^i\beta\} = 1$$

Donc

$$\text{Prob}\{\epsilon_i = -x^i\beta\} = 1 - \text{Prob}\{\epsilon_i = 1-x^i\beta\}$$

En remplaçant cette dernière expression dans celle de l'espérance mathématique de ϵ_i , il vient :

$$\text{Prob}\{\epsilon_i = -x^i\beta\} = 1 - x^i\beta$$

et

$$\text{Prob}\{\epsilon_i = 1-x^i\beta\} = x^i\beta$$

Dans le cas discret, la variance de l'erreur aléatoire s'exprime de la façon suivante :

$$\text{Var}(\epsilon_i) = E(\epsilon_i^2) = \text{Prob}\{\epsilon_i = x^i\beta\} (-x^i\beta)^2 + \text{Prob}\{\epsilon_i = 1-x^i\beta\} (1-x^i\beta)^2$$

Soit en remplaçant les probabilités par leurs valeurs :

$$E(\epsilon_i^2) = x^i\beta(1-x^i\beta) = \sigma^2_{i_i}$$

De même

$$E(\epsilon_j^2) = x_j^j \beta (1 - x_j^j \beta) = \sigma^2_{jj}$$

Les erreurs aléatoire n'étant pas homoscedastiques, les estimateurs des paramètres inconnus du modèle, obtenus par la méthode des moindres carrés ordinaires bien que linéaires et centrés sont néanmoins inefficients. Négliger l'hétéroscedasticité des erreurs reviendrait à sous-estimer les vraies variances des estimateurs des paramètres inconnus et donc à biaiser les tests vers l'acceptation d'hypothèse.

I.3. ESTIMATION PAR LES MOINDRES CARRÉS GÉNÉRALISÉS

Pour pallier le problème de l'hétéroscedasticité des erreurs, il est alors nécessaire d'estimer les paramètres inconnus du modèle par la méthode des moindres carrés généralisés ou méthode de Aitken. Celle-ci consiste à transformer le modèle initial de telle sorte qu'ex post les erreurs deviennent homoscedastiques. Pour ce faire, il suffit alors de pondérer chaque observation par l'inverse de l'écart-type de l'erreur correspondante. Le modèle à estimer est de la forme :

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \epsilon_i \quad i = 1, \dots, n$$

avec $E(\epsilon_i) = 0$

$$\begin{aligned} E(\epsilon_i \epsilon_j) &= \sigma^2_{ii} = \sigma^2 \lambda_i^2 && \text{pour } i = j \\ &= 0 && \text{pour } i \neq j \end{aligned}$$

En divisant par l'écart-type de l'erreur il vient :

$$\frac{y_i}{\lambda_i} = \frac{\beta_0}{\lambda_i} + \beta_1 \frac{x_{1i}}{\lambda_i} + \dots + \beta_k \frac{x_{ki}}{\lambda_i} + \frac{\epsilon_i}{\lambda_i} \quad i = 1, \dots, n$$

De ce fait, l'erreur transformée, $\frac{\epsilon_i}{\lambda_i}$, est bien homoscedastique.

En effet,

$$\text{Var}\left(\frac{\epsilon_i}{\lambda_i}\right) = \frac{1}{\lambda_i^2} \text{Var}(\epsilon_i) = \frac{\sigma^2 \lambda_i^2}{\lambda_i^2} = \sigma^2 \quad \forall i$$

L'application de la méthode des moindres carrés ordinaires sur le modèle transformé permet alors de découvrir les estimateurs des moindres carrés généralisés du modèle initial. Ces estimateurs, linéaires par rapport à la variable endogène y_i et centrés, sont à variance minimale. Par conséquent ce sont des estimateurs BLUE.

Dans le cas qui nous préoccupe, les variances des erreurs étant inconnues, la méthode de Aitken pure n'est pas réalisable. Il est donc indispensable de donner auparavant, une estimation convergente de chacune des erreurs. Mac Gillivray ("Econometrica", Vol.38, n°5, 1970, pp.775-776) suggère de prendre comme estimateur de la variance l'expression suivante :

$$\widehat{\text{Var}}(\varepsilon_i) = \widehat{y}_i(1-\widehat{y}_i)$$

où \widehat{y}_i est la valeur calculée du modèle.

$$y_i = x_i^1 \beta + \varepsilon_i$$

estimé par la méthode des moindres carrés ordinaires.

Cependant, il n'est pas exclu que certaines variances soient négatives. Il est alors possible de tourner la difficulté en choisissant de prendre

$$\widehat{\text{Var}}(\varepsilon_i) = |\widehat{y}_i(1-\widehat{y}_i)|$$

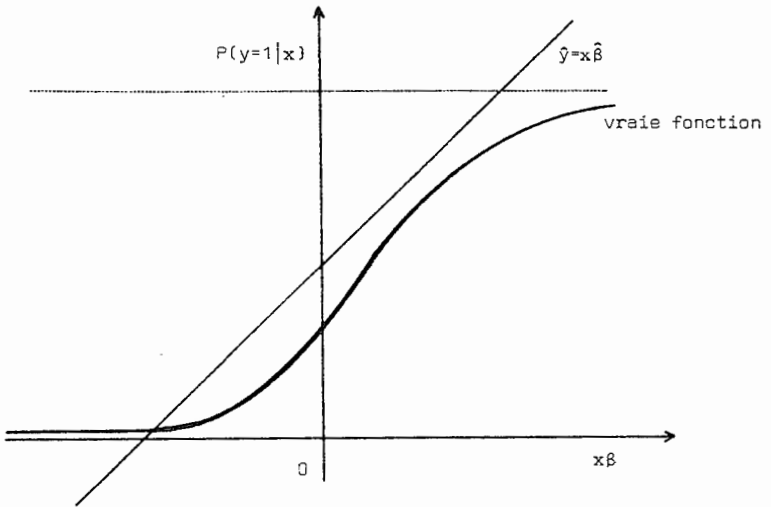
Bien que l'estimation de la fonction de probabilité linéaire par la méthode des moindres carrés généralisés constitue une amélioration par rapport à son estimation par les moindres carrés ordinaires, il n'en reste pas moins qu'un problème important subsiste. En effet, l'hypothèse de normalité des erreurs habituellement postulée pour juger de la significativité des variables n'est plus acceptable dans le cas d'un modèle à variable dépendante dichotomique. La raison en est que l'erreur aléatoire prend ses valeurs dans l'intervalle $[-x_i^1 \beta, 1-x_i^1 \beta]$ et non pas dans l'intervalle $]-\infty, +\infty[$.

De ce fait, les tests de significativité des variables, du type tests de Student, sont nécessairement biaisés.

Hormis ceci, un problème important subsiste, à savoir que le modèle linéaire est inadéquat pour représenter une probabilité. En effet, la valeur calculée

$$\hat{y}_i = x_i^t \hat{\beta}$$

n'est rien d'autre que l'estimation de la probabilité conditionnelle de réalisation de l'évènement expliquée. La caractéristique d'une probabilité est d'être comprise entre zéro et un. Or avec le modèle linéaire, rien ne nous assure que la valeur calculée restera comprise dans cet intervalle.



Ceci est un handicap sérieux lorsqu'il faut établir des prédictions et plutôt que d'estimer une fonction de probabilité linéaire il est préférable d'estimer une fonction non linéaire prenant ses valeurs dans l'intervalle $[0,1]$ et dans laquelle y_i est une fonction non décroissante de $x_i^t \beta$.

I.4. EXEMPLE D'APPLICATION

Les résultats présentés ci-après sont obtenus à partir d'un échantillon de 214 étudiants inscrits pour la première fois à l'U.E.R. de Médecine de l'Université de Dijon, au début de l'année scolaire 1974-75¹. L'évènement expliqué est la réussite d'un étudiant à l'examen de fin d'année d'études. Les variables exogènes supposées déterminer ce phénomène² ainsi que le signe attendu des coefficients de régression (un signe positif augmente la probabilité de réussite, un signe négatif augmente la probabilité d'échec) sont les suivantes :

| | $\frac{\partial P}{\partial x}$ | anticipé |
|--|---------------------------------|----------|
| 1. Taille de la commune de résidence des parents | | ? |
| 2. Revenus mensuels des parents | | + |
| 3. Age de l'étudiant | | - |
| 4. Résultat à un test d'aptitude logique | | + |
| 5. Résultat à un test de personnalité | | - |
| 6. Moyenne à l'écrit du baccalauréat | | + |
| 7. Etudes précédentes 1 si l'étudiant était déjà dans le supérieur en 74/75 0 sinon | | - |
| 8. Origine du secondaire 1 si l'étudiant a effectué ses études secondaires dans un établissement public, 0 sinon | | + |
| 9. Baccalauréat scientifique 1 si l'étudiant possède un bac. série C 0 sinon | | + |
| 10. Baccalauréat non scientifique 1 si l'étudiant possède un bac série A,B,F ou G, 0 sinon. | | - |

¹ Cet échantillon fait partie d'une étude réalisée par l'Institut de Recherche sur l'Economie de l'Education et financé par le Service d'Etudes et d'Informations Statistiques du Ministère des Universités.

² Voir à ce propos, A. MINGAT, "La première année d'études, la réussite, l'abandon, l'échec". Cahier de l'IREDU n° 23.

Les résultats obtenus par les moindres carrés ordinaires sont les suivants :

| VARIABLE | COEFFICIENT |
|--------------------------|-------------|
| Taille de la commune | - 0,043* |
| Revenus des parents/1000 | 0,013 |
| Age/10 | - 0,456 |
| Test logique | 0,038 |
| Test de personnalité/10 | - 0,030*** |
| Moyenne à l'écrit du bac | 0,078*** |
| Etudes précédentes | 0,232* |
| Origine du secondaire | 0,050 |
| Baccalauréat C | 0,233*** |
| Baccalauréat A, B, F, G | - 0,099 |
| Constante | 0,207 |
| R ² = 0,30 | |

TABLEAU I.1. : ESTIMATION DE LA FONCTION DE PROBABILITE LINEAIRE PAR LES MOINDRES CARRÉS ORDINAIRES.

Les tests de significativité des variables sont construits sous l'hypothèse de normalité des erreurs. Les seuils retenus sont les suivants :

* = 10 % ** = 5 % *** = 1 %

Posons-nous la question de savoir s'il est licite d'estimer ce modèle par les moindres carrés ordinaires et par conséquent si nous pouvons accepter l'hypothèse d'homoscédasticité des erreurs. Pour ce faire, il suffit de construire un test de non-homoscédasticité¹ basé sur l'estimation du modèle pour deux sous-populations. La comparaison du rapport entre la somme des carrés des résidus du modèle estimé par les moindres carrés ordinaires sur la première sous-population et la somme des carrés des résidus du modèle estimé par la même méthode sur la seconde sous-population avec un F de Fisher théorique, indique que nous ne pouvons pas écarter l'hypothèse que les erreurs soient en fait hétéroscédastiques.

¹ Pour une description théorique de ce test, voir Theil, H. "Principles of Econometrics", Wiley, New-York. 1971 - pp.196-197.

En effet, nous obtenons :

$$\frac{SS \text{ de la 1ère sous-population}}{SS \text{ de la 2ème sous-population}} = 2,43$$

alors que $F(n_1 - 11, n_2 - 11) = 1,39$ (avec n_1 et n_2 les effectifs de chacune des deux sous-populations).

L'estimation du modèle linéaire de réussite par la méthode de Aitken réalisable est alors la suivante :

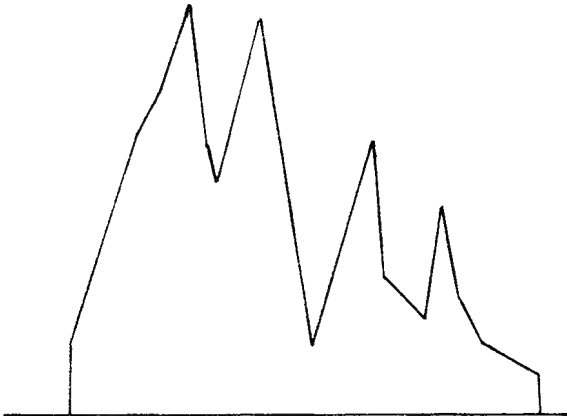
| VARIABLE | COEFFICIENT |
|--------------------------|--------------|
| Taille de la commune | - 0,028 |
| Revenus des parents/1000 | 0,013* |
| Age/10 | - 0,430** |
| Test logique | 0,001 |
| Test de personnalité/10 | - 0,030*** |
| Moyenne à l'écrit du bac | 0,078*** |
| Etudes précédentes | 0,209* |
| Origine du secondaire | 0,074 |
| Baccalauréat C | 0,210*** |
| Baccalauréat A, B, F, G | - 0,129*** |
| Constante | 0,174 |
| | $R^2 = 0,29$ |

TABLEAU I.2.: ESTIMATION DE LA FONCTION DE PROBABILITE LINEAIRE PAR LES MOINDRES CARRES GENERALISES.

Alors qu'à l'issue de l'estimation du modèle par les moindres carrés ordinaires, seulement cinq variables sur les dix initialement retenues sont significatives à un seuil au moins égal à 10 %, avec la méthode des moindres carrés généralisés, trois variables supplémentaires sont significatives. Etant donné que l'hypothèse de non normalité des erreurs (voir histogramme des résidus) défavorise aussi bien les tests dans l'une ou l'autre méthode, il faut attribuer cette supériorité de significativité à une plus grande efficacité des estimateurs des moindres carrés généralisés.

Il suffit pour s'en convaincre de comparer les variances des estimateurs obtenus par l'une et l'autre méthode :

| | VARIANCE DES ESTIMATEURS | |
|--------------------------|--------------------------|------------------|
| | DES M C O | DES M C G |
| Taille de la commune | $0,669.10^{-3}$ | $0,371.10^{-3}$ |
| Revenus des parents | $0,936.10^{-10}$ | $0,439.10^{-10}$ |
| Age | $0,806.10^{-5}$ | $0,352.10^{-5}$ |
| Test logique | $0,442.10^{-4}$ | $0,144.10^{-4}$ |
| Test de personnalité | $1,231.10^{-8}$ | $0,709.10^{-8}$ |
| Moyenne à l'écrit du bac | $0,241.10^{-5}$ | $0,132.10^{-5}$ |
| Etudes précédentes | $0,158.10^{-1}$ | $0,133.10^{-1}$ |
| Origine du secondaire | $0,568.10^{-2}$ | $0,275.10^{-2}$ |
| Baccalauréat C | $0,358.10^{-2}$ | $0,209.10^{-2}$ |
| Baccalauréat A,B,F,G | $1,59.10^{-2}$ | $0,387.10^{-2}$ |



GRAPHIQUE I.3. = HISTOGRAMME DES RESIDUS.

Nous donnons ci-après les prédictions de la probabilité de réussite pour un sous-échantillon aléatoire de 40 étudiants.

| N° d'ob- servation | valeur observée | Prédiction | | N° d'ob- servation | valeur observée | Prédiction | |
|-----------------------|--------------------|------------|--------|-----------------------|--------------------|------------|--------|
| | | M C O | M C G | | | M C O | M C G |
| 159 | 1 | 0,680 | 0,665 | 67 | 0 | 0,229 | 0,242 |
| 189 | 0 | -0,103 | -0,031 | 149 | 0 | -0,107 | -0,099 |
| 76 | 0 | -0,016 | -0,014 | 163 | 1 | 0,582 | 0,583 |
| 167 | 1 | 0,475 | 0,449 | 47 | 1 | 0,291 | 0,300 |
| 7 | 0 | -0,033 | -0,036 | 106 | 0 | 0,313 | 0,289 |
| 34 | 1 | 0,958 | 0,941 | 102 | 0 | 0,446 | 0,447 |
| 81 | 0 | 0,502 | 0,461 | 197 | 1 | 0,611 | 0,614 |
| 20 | 1 | 0,668 | 0,647 | 33 | 0 | 0,184 | 0,198 |
| 66 | 1 | 0,657 | 0,646 | 29 | 1 | 0,443 | 0,452 |
| 127 | 0 | 0,265 | 0,278 | 178 | 0 | 0,612 | 0,608 |
| 5 | 0 | -0,043 | -0,025 | 2 | 0 | 0,076 | 0,082 |
| 212 | 0 | 0,190 | 0,200 | 158 | 0 | 0,147 | 0,189 |
| 44 | 0 | 0,318 | 0,359 | 103 | 0 | 0,584 | 0,573 |
| 184 | 1 | 0,458 | 0,449 | 91 | 1 | 0,626 | 0,592 |
| 121 | 0 | 0,021 | 0,094 | 179 | 0 | 0,028 | 0,035 |
| 97 | 1 | 0,185 | 0,198 | 93 | 0 | 0,528 | 0,585 |
| 118 | 0 | 0,000 | -0,002 | 75 | 1 | 0,432 | 0,414 |
| 111 | 1 | 1,055 | 1,017 | 101 | 0 | -0,011 | -0,023 |
| 21 | 1 | 0,046 | 0,085 | 169 | 0 | -0,047 | -0,023 |
| 162 | 0 | -0,318 | -0,312 | 12 | 0 | 0,182 | 0,168 |

TABLEAU I.4.: PREDICTIONS DE LA PROBABILITE LINEAIRE PAR LES MOINDRES CARRES ORDINAIRES ET LES MOINDRES CARRES GENERALISES.

Si nous calculons pour chaque modèle, la distance entre les valeurs observées et les valeurs prédites ci-dessus, nous obtenons :

prédiction des moindres carrés ordinaires : 6,11
 prédiction des moindres carrés généralisés: 6,06

Ainsi pour l'échantillon considéré, et toutes choses égales par ailleurs, les prédictions obtenues par les moindres carrés généralisés sont meilleures que celles obtenues par les moindres carrés ordinaires. Il n'en reste pas moins qu'un problème important subsiste, à savoir que la forme linéaire est inadéquate pour représenter une probabilité. Il en est pour preuve les prédictions figurant au tableau précédent et dans lequel certains individus ont une probabilité de réussite négative, voire supérieure à un.

CHAPITRE II

LES TECHNIQUES D'ESTIMATION DE DONNEES GROUPEES

Le modèle de régression linéaire présente de nombreux inconvénients parmi lesquels celui de ne pas imposer à la valeur calculée du modèle d'être comprise dans l'intervalle $[0,1]$. Un moyen d'éviter cet écueil consiste alors à représenter le phénomène expliqué par une forme non linéaire variant dans l'intervalle $[0,1]$, puis par transformation du modèle à se ramener à l'expression linéaire usuelle. Cette procédure, calquée sur les pratiques des biologistes, nécessite au préalable la définition de données groupées.

Plusieurs méthodes sont alors possibles, mais toutes présentent l'inconvénient de réduire considérablement l'information disponible, d'une part, d'imposer l'arbitraire à l'analyste d'autre part.

II.1. PROBIT ANALYSIS¹

Cette méthode, comme la plupart des autres présentées ci-après, a été utilisée en premier lieu par les biologistes pour expliquer les effets d'un poison ou d'une drogue sur une population d'animaux ou de plantes. Les réponses d'individus à un quelconque stimulus, qu'examinent ces chercheurs sont comparables à certaines réactions d'agents économiques. Par exemple, pour chaque famille, il existe un certain niveau de revenu en-deçà duquel elle ne possède pas d'automobile et au-delà duquel elle en possède une. Ce niveau qu'un biologiste appellerait tolérance est une variable aléatoire et peut donc être caractérisé par une fonction de densité de probabilité. Si un revenu x_0 est attribué à une famille et si $f(x)$ représente la densité de la tolérance alors la probabilité de réalisation de l'évènement pour cette famille est donnée par :

$$P = \int_0^{x_0} f(x)dx$$

Dans le cas précis de la Probit Analysis, supposons une population d'animaux sur laquelle nous étudions les effets de différentes doses de poison. Cette population est répartie en G groupes n_i , $i = 1, \dots, G$. A chaque animal d'un groupe n_i nous administrons une dose t_i de poison. Notons $y_{ij} = 1$ si l'animal j du groupe i meurt à la suite de l'injection t_i de drogue, et $y_{ij} = 0$ sinon. Soit $\text{Prob}\{y_{ij}=1\}$ la probabilité qu'un animal du groupe i ne survive pas et $(a+bt_i)$ le niveau de poison à partir duquel l'animal meurt. Si nous représentons cette probabilité par une fonction non décroissante de t_i , nous avons :

$$\text{Prob}\{y_{ij}=1\} = F(a+bt_i)$$

¹ Le terme "Probit" proposé par C.I. BLISS (1934) est la contraction de "Probability Unit".

où F est une fonction de répartition de sorte que p_i est nécessairement comprise dans l'intervalle $[0,1]$.

De façon plus générale, soit x^i_β le niveau à partir duquel l'évènement expliqué se réalise. Le modèle s'écrit alors

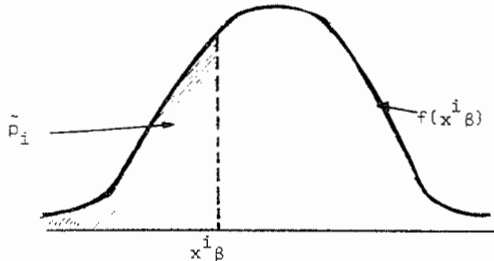
$$\text{Prob}\{y_{ij}=1\} = F(x^i_\beta).$$

La Probit Analysis consiste à utiliser pour F la fonction de répartition d'une variable aléatoire normale centrée réduite. Elle nécessite au préalable la définition de données groupées afin de remplacer la variable dichotomique expliquée par sa fréquence d'apparition dans chaque groupe initialement défini.

Soit \tilde{p}_i , l'estimateur de la probabilité de réalisation de l'évènement expliqué à l'intérieur de chaque groupe i , nous avons donc

$$\tilde{p}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \quad i = 1, \dots, G$$

Soit $f(x^i_\beta)$ la fonction de densité de probabilité du niveau de réaction de l'individu i ; \tilde{p}_i peut alors être représenté de la façon suivante :



Dans le cas de l'utilisation d'une variable aléatoire normale centrée réduite nous avons donc :

$$\tilde{p}_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x^i \beta} \exp \left\{ -\frac{1}{2} u^2 \right\} du$$

En posant $Z_i = x^i \beta$, il vient alors :

$$\tilde{p}_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z_i} \exp \left\{ -\frac{1}{2} u^2 \right\} du$$

Ainsi la Probit Analysis consiste à remplacer une variable contrainte par une variable Z_i prenant ses valeurs dans l'intervalle $]-\infty, +\infty[$. Désormais, il est possible d'adopter le modèle de régression linéaire, à savoir :

$$Z_i = x^i \beta + \varepsilon_i \quad i = 1, \dots, G$$

avec les hypothèses

$$E(\varepsilon_i) = 0 \quad \forall i$$

$$E(\varepsilon_i \varepsilon_j) = \sigma^2 \text{ pour } i = j \\ = 0 \text{ sinon}$$

et d'estimer ce modèle par la méthode des moindres carrés ordinaires. Toutefois il est nécessaire qu'il existe dans chaque groupe d'individus préalablement défini au moins un individu pour lequel l'évènement expliqué se réalise, sinon Z_i n'est pas déterminé. Outre cette restriction, la Probit Analysis présente deux inconvénients majeurs. Le calcul de l'intégrale ci-dessus n'est pas simple et nécessite le recours à un processus numérique. Quand bien même cela ne serait pas, l'analyse des phénomènes économiques n'est pas analogue à l'analyse des phénomènes biologiques et en aucun cas l'économetre ne dispose de données groupées. Or mis à part l'arbitraire que cette technique impose, il est extrêmement difficile de grouper les individus selon les valeurs de leurs variables et ceci d'autant plus que leur nombre est élevé.

II.2. LOGIT ANALYSIS¹

Alors que la Probit Analysis utilise la fonction de répartition d'une variable aléatoire normale centrée réduite pour contraindre la valeur calculée de la variable dépendante dans l'intervalle $[0,1]$, la Logit Analysis utilise quant à elle la fonction de distribution logistique standardisée. Cette fonction, connue également sous le nom de loi de Verhulst, s'écrit de la façon suivante :

$$y = \frac{1}{1+e^{-bx}} \quad -\infty < x < +\infty$$

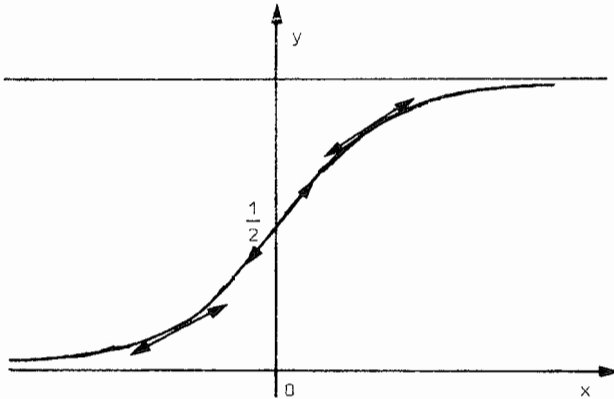
Employée aussi bien dans le cas de données groupées que dans le cas de données individuelles (cf. chapitre III : L'analyse logistique des données individuelles), cette fonction mérite une attention particulière en raison de sa simplicité, comparativement aux autres fonctions utilisées dans l'estimation de modèles à variable dépendante dichotomique.

Cette fonction possède les propriétés suivantes : lorsque $x = 0$, $y = \frac{1}{2}$; lorsque $x = -\infty$, $y = 0$ et lorsque $x = \infty$, $y = 1$. Cette fonction admet donc 0 et 1 comme asymptote. Le point $x = 0$ et $y = \frac{1}{2}$ est centre de symétrie. La pente en chaque point est égale à

$$\frac{dy}{dx} = by(1 - y)$$

Au point $x = 0$, la pente atteint sa valeur maximale, à savoir $\frac{b}{4}$. Etant donné que la dérivée seconde de y par rapport à x s'annule en ce point, celui-ci est également point d'inflexion. La représentation graphique de cette fonction est alors la suivante :

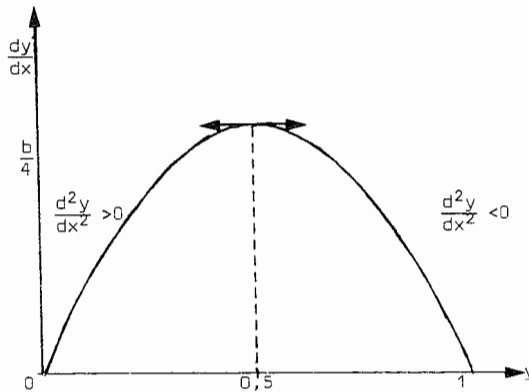
¹ Le terme "Logit" proposé par J. BERKSON (1944) est la contraction de "Logistic Unit".



La dérivée de y par rapport à x , à savoir

$$\frac{dy}{dx} = by(1-y)$$

est quant à elle une parabole dont le maximum $\frac{b}{4}$ est atteint $y = \frac{1}{2}$. La représentation graphique de cette dérivée est alors la suivante :



Pour reprendre la terminologie utilisée dans l'exposé de la Probit Analysis, rappelons que le modèle à estimer est de la forme :

$$\text{Prob}\{y_{ij}=1\} = F(x^i\beta)$$

Si nous adoptons la fonction logistique pour représenter le phénomène étudié, nous avons alors :

$$\text{Prob}\{y_{ij}=1\} = \frac{1}{1+e^{-x_{ij}^i \beta}}$$

Nous pouvons donc écrire

$$1 - \text{Prob}\{y_{ij}=1\} = 1 - \frac{1}{1+e^{-x_{ij}^i \beta}} = \frac{e^{-x_{ij}^i \beta}}{1+e^{-x_{ij}^i \beta}}$$

$$\text{Soit encore} \quad \frac{\text{Prob}\{y_{ij}=1\}}{1 - \text{Prob}\{y_{ij}=1\}} = e^{x_{ij}^i \beta}$$

En prenant le logarithme des deux membres, nous avons :

$$\ln \frac{\text{Prob}\{y_{ij}=1\}}{1 - \text{Prob}\{y_{ij}=1\}} = x_{ij}^i \beta$$

Ainsi comme dans la Probit Analysis, nous sommes ramenés au modèle linéaire classique. Si nous estimons la probabilité de réalisation de l'évènement expliqué par sa fréquence d'apparition dans chaque groupe d'individus préalablement défini, nous avons alors :

$$\ln \frac{\tilde{p}_i}{1 - \tilde{p}_i} = x_i^i \beta$$

$$\text{avec } \tilde{p}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \quad i = 1, \dots, G$$

Là encore il est possible d'utiliser les moindres carrés ordinaires pour découvrir les estimateurs des paramètres inconnus du modèle. Cette méthode, séduisante par sa simplicité nécessite l'usage de données groupées. Bien qu'il soit toujours possible de partitionner les individus d'un échantillon selon les valeurs des variables qui leur sont affectées, il est cependant préférable de développer des méthodes utilisant des données individuelles de manière à éviter l'inconvénient des groupes vides d'une part, à ne pas réduire

l'information disponible lorsque l'échantillon des observations est réduit, d'autre part.

A ces problèmes s'ajoute celui de l'estimation du modèle. En effet la Logit Analysis, telle qu'elle a été présentée ci-dessus, repose sur la transformation d'un modèle non linéaire en un modèle linéaire. Or, cette transformation ne s'applique pas seulement aux variables endogène et exogènes mais également au terme d'erreur aléatoire, si bien que celui-ci ne satisfait plus les hypothèses usuelles. En effet, nous avons

$$\tilde{p}_i = \frac{1}{1+e^{-x^i\beta}} + \epsilon_i$$

En opérant la transformation décrite ci-dessus, nous avons en réalité le modèle suivant :

$$\frac{\tilde{p}_i}{1-\tilde{p}_i} = \frac{1 + \epsilon_i(1+e^{-x^i\beta})}{e^{-x^i\beta} - \epsilon_i(1+e^{x^i\beta})}$$

Soit en prenant le logarithme des deux membres

$$\ln \frac{\tilde{p}_i}{1-\tilde{p}_i} = \ln [1+\epsilon_i(1+e^{-x^i\beta})] - \ln [e^{-x^i\beta} - \epsilon_i(1+e^{x^i\beta})]$$

et non pas

$$\ln \frac{\tilde{p}_i}{1-\tilde{p}_i} = x^i\beta + \epsilon_i$$

En toute logique, ce modèle ne constitue en aucun cas le transformé du modèle initial. Son estimation revient à ne pas tenir compte des caractéristiques de l'erreur aléatoire supposée représenter les facteurs omis lors de l'explication de la fréquence d'apparition de l'évènement considéré. Or ce faisant, l'estimation de la probabilité est nécessairement biaisée. Pour pallier cet inconvénient, il serait alors préférable de ne pas transformer le modèle, ce qui permettrait non seulement de résoudre le problème du groupement d'individus, mais également celui des tests de significativité des variables supposées déterminer le phénomène expliqué.

II.3. TRANSFORMATIONS DE LA FONCTION DE PROBABILITE LINEAIRE

La différence essentielle entre les méthodes de transformation de la fonction de probabilité linéaire et les méthodes présentées ci-dessus, réside dans l'introduction ex post de la non linéarité d'une part, dans la nature des données traitées d'autre part, puisque ces méthodes ne nécessitent pas la définition préalable de groupes d'individus selon les valeurs des variables exogènes qui leur sont affectées mais la définition des groupes d'individus selon la valeur calculée de leur probabilité de réalisation de l'évènement expliqué ou selon le caractère dichotomique de la variable endogène étudiée.

La méthode dite de transformation logistique de la fonction de probabilité linéaire consiste à ranger en classes les valeurs prédites obtenues lors de l'estimation de la forme linéaire

$$y_i = x_i^1 \beta + \epsilon_i$$

et à calculer pour chacune d'elle l'expression

$$\hat{L} = \ln \left(\frac{\bar{y}}{1-\bar{y}} \right)$$

où \bar{y} représente la moyenne des prédictions de chaque intervalle, obtenue par la méthode des moindres carrés généralisés, appliquée au modèle

$$y_i = x_i^1 \beta + \epsilon_i$$

La régression linéaire de cette expression sur les centres de classes permet de mettre en relation \hat{L} avec \hat{y} et donc indirectement avec le vecteur des variables exogènes. On exprime alors la probabilité de l'évènement considéré, d'un individu i , de la façon suivante :

$$\widehat{\text{Prob}} \{y_i=1\} = \frac{1}{1+e^{-\hat{L}_i}}$$

Si cette méthode permet de ne plus obtenir des valeurs prédites négatives ou supérieures à un, elle n'est cependant pas totalement satisfaisante car elle est basée sur l'estimation de la fonction de probabilité linéaire par les moindres carrés généralisés, or les tests de significativité des variables sont nécessairement biaisés. De plus, elle nécessite la définition arbitraire d'intervalles pour les valeurs \hat{y}_i .

La méthode dite de transformation de Warner de la fonction de probabilité linéaire consiste à estimer la probabilité de réalisation de l'évènement expliqué par :

$$\widehat{\text{Prob}}\{y_i=1\} = \frac{e^{\hat{D}(x^1)}}{1+e^{\hat{D}(x^1)}}$$

où \hat{D} est l'estimateur de la fonction discriminante. Cette fonction est en relation avec \hat{y} de la façon suivante :

$$\hat{D}(x^1) = K(\hat{y}_1 - \hat{\beta}_0)$$

$$\text{avec } K = \frac{N_1 + N_2 - 2}{\frac{N_1 N_2}{N_1 + N_2} - \frac{N_1 N_2}{N_1 + N_2} (\bar{x}^1 + \bar{x}^2) \hat{\beta}}$$

et $\hat{\beta}_0$ l'estimateur des moindres carrés généralisés de la constante de la fonction de probabilité linéaire.

N_1 étant le nombre d'individus ayant 1 pour valeur de la variable endogène, N_2 le nombre d'individus ayant la valeur 0, \bar{x}^1 le vecteur des moyennes des variables exogènes du groupe d'individus ayant 1 pour valeur de la variable endogène, \bar{x}^2 le vecteur des moyennes des variables exogènes du groupe d'individus ayant 0 pour valeur de la variable endogène et $\hat{\beta}$ le vecteur des estimateurs des moindres carrés généralisés des paramètres inconnus de la fonction de probabilité linéaire.

Les critiques adressées à la méthode de transformation logistique de la fonction de probabilité linéaire s'appliquent également à la méthode de transformation de Warner. Cette dernière présente

toutefois un léger avantage par rapport à la précédente, à savoir qu'elle ne nécessite pas la création arbitraire d'intervalles pour les valeurs de \hat{y}_i . Malgré tout, elle aussi introduit artificiellement la non linéarité sans tenir compte de ses effets sur les estimateurs des paramètres inconnus du modèle d'une part, sur la significativité des variables supposées déterminer le phénomène expliqué d'autre part.

II.4. EXEMPLE D'APPLICATION

Les résultats illustrant la Probit Analysis et la Logit Analysis sont obtenus à partir de renseignements départementaux sur la scolarisation des étudiants en Droit durant l'année scolaire 1960-61.¹

Soit N le nombre total d'individus en âge d'être scolarisés. Cette population est répartie en 90 groupes n_i , $i=1, \dots, 90$, c'est-à-dire en autant de groupes qu'il y a de départements. Notons $y_{ij} = 1$ si un individu j du département i suit un enseignement juridique universitaire, et $y_{ij} = 0$ sinon. Soit \hat{p}_i l'estimation de la probabilité qu'un individu du département i soit inscrit dans cette discipline. Nous avons alors :

$$\hat{p}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

L'expression $\sum_{j=1}^{n_i} y_{ij}$ n'est rien d'autre que le nombre d'étudiants juristes du département. Le rapport \hat{p}_i , dans la mesure où n_i est judicieusement choisi, est alors équivalent au taux de scolarisation en Droit du département i .

Les variables exogènes supposées déterminer ce phénomène² ainsi que

¹ Cet échantillon fait partie d'une étude réalisée par l'IREDU et financée par le C.N.R.S.

² Voir à ce propos G.LASSIBILLE, A.MINGAT, J.PERROT "Les effets de la modification de la carte universitaire - 1960-1975". Cahier de l'IREDU n°25.

le signe attendu des coefficients de régression sont les suivants :

- | | |
|---|---------------------------------|
| | $\frac{\partial F}{\partial x}$ |
| 1 - Présence dans le département d'un gros établissement universitaire de Droit (PGED) | + |
| 1 s'il existe un tel établissement 0 sinon. | |
| 2 - Présence dans le département d'un petit établissement universitaire de Droit (PPED) | + |
| 1 s'il existe un tel établissement 0 sinon | |
| 3 - Absence dans le département d'un établissement universitaire de Droit, mais présence d'un établissement universitaire d'une autre discipline (ADP \bar{D}) | - |
| 1 si oui 0 sinon | |

Les résultats obtenus par les moindres carrés ordinaires sur le modèle

$$\begin{cases} \tilde{p}_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z_i} \exp\left\{-\frac{1}{2} u^2\right\} du \\ Z_i = x^i \beta + \epsilon_i \end{cases}$$

et sur le modèle

$$\ln \frac{\tilde{p}_i}{1-\tilde{p}_i} = x^i \beta + \epsilon_i$$

sont les suivants :

| VARIABLE | COEFFICIENT PROBIT ANALYSIS | COEFFICIENT LOGIT ANALYSIS |
|-----------------|--------------------------------|-------------------------------|
| P G E D | 0,258*** | 0,537*** |
| P P E D | 0,123* | 0,253* |
| A D P \bar{D} | -0,041 | -0,082 |
| CONSTANTE | -1,506 | -2,658 |
| | R ² = 0,23 | R ² = 0,23 |

TABLEAU II-1 : PROBIT ANALYSIS ET LOGIT ANALYSIS DES TAUX DE SCOLARISATION DEPARTEMENTAUX EN DROIT (1960-61)

Les tests de significativité des variables sont construits sur l'hypothèse de normalité des erreurs. Les seuils retenus sont les suivants :

* = 10 % ** = 5 % *** = 1 %

Au vu des résultats il est à remarquer que les effets marginaux obtenus par la Logit Analysis sont environ deux fois plus importants que ceux obtenus par la Probit Analysis. Toutefois, en raison de la même différence entre les constantes des deux modèles, les prédictions fournies par l'une et l'autre méthode sont sensiblement équivalentes. Pour s'en convaincre, il suffit d'examiner les estimations du taux de scolarisation en Droit pour les quatre départements hypothétiques suivants¹:

| | <u>Probit Analysis</u> | <u>Logit Analysis</u> |
|---|----------------------------|---------------------------|
| A - Département totalement dépourvu d'établissement universitaire de Droit | 0,656 | 0,655 |
| B - Département pourvu d'un gros établissement universitaire de Droit | 1,057 | 1,070 |
| C - Département pourvu d'un petit établissement universitaire de Droit | 0,838 | 0,827 |
| D - Département non pourvu d'un établissement universitaire de Droit, mais pourvu d'un établissement universitaire dans une autre discipline. | 0,606 | 0,606 |

Les résultats illustrant la transformation logistique et la transformation de Warner de la fonction de probabilité linéaire sont obtenus à partir de l'estimation par les moindres carrés généralisés du modèle linéaire de réussite présenté au chapitre I. (Tableau I.2)

¹ Les taux sont exprimés en pourcentage.

En ce qui concerne la transformation logistique de la fonction de probabilité linéaire, nous obtenons l'expression suivante :

$$\hat{L}_i = -2,798 + 5,557 \hat{y}_i$$

La probabilité de réussite d'un individu i , s'exprime ainsi :

$$\widehat{\text{Prob}}\{y_i=1\} = \frac{1}{1+e^{-\hat{L}_i}}$$

Par exemple, si nous considérons le premier individu du sous-échantillon aléatoire présenté page 16, il faut pour donner une estimation de sa probabilité de réussite "révisée" par cette méthode, calculer

$$\hat{L}_i = -2,798 + (5,557 \times 0,665) = 0,897$$

puis

$$\widehat{\text{Prob}}\{y_i=1\} = \frac{1}{1+e^{-0,897}} = 0,710$$

En ce qui concerne la transformation de Warner de la fonction de probabilité linéaire, nous obtenons l'expression suivante :

$$\hat{D}(x^i) = -1,200 + 6,920 \hat{y}_i$$

Le premier individu du sous échantillon aléatoire a donc

$$\hat{D}(x^i) = -1,200 + (6,920 \times 0,665) = 3,332$$

Sa probabilité de réussite est alors égale à

$$\widehat{\text{Prob}}\{y_i=1\} = \frac{e^{3,332}}{1+e^{3,332}} = 0,965$$

Le tableau ci-dessous donne l'estimation de la probabilité de réussite révisée par l'une et l'autre méthode pour l'ensemble du sous échantillon.

| N° d'observation | Valeur observée | Prédiction transformation Logistique | Prédiction transformation Warner | N° d'observation | Valeur observée | Prédiction transformation Logistique | Prédiction transformation Warner |
|------------------|-----------------|--------------------------------------|----------------------------------|------------------|-----------------|--------------------------------------|----------------------------------|
| 159 | 1 | 0,710 | 0,965 | 67 | 0 | 0,169 | 0,617 |
| 189 | 0 | 0,048 | 0,194 | 149 | 0 | 0,033 | 0,131 |
| 76 | 0 | 0,053 | 0,214 | 163 | 1 | 0,608 | 0,944 |
| 167 | 1 | 0,424 | 0,871 | 47 | 1 | 0,243 | 0,705 |
| 7 | 0 | 0,047 | 0,226 | 106 | 0 | 0,232 | 0,690 |
| 34 | 1 | 0,919 | 0,995 | 102 | 0 | 0,422 | 0,669 |
| 81 | 0 | 0,441 | 0,879 | 197 | 1 | 0,648 | 0,954 |
| 20 | 1 | 0,689 | 0,963 | 33 | 0 | 0,154 | 0,543 |
| 56 | 1 | 0,688 | 0,963 | 29 | 1 | 0,428 | 0,873 |
| 127 | 0 | 0,222 | 0,673 | 178 | 0 | 0,641 | 0,952 |
| 5 | 0 | 0,050 | 0,201 | 2 | 0 | 0,087 | 0,347 |
| 212 | 0 | 0,156 | 0,546 | 158 | 0 | 0,146 | 0,526 |
| 44 | 0 | 0,309 | 0,783 | 103 | 0 | 0,595 | 0,940 |
| 184 | 1 | 0,424 | 0,780 | 91 | 1 | 0,620 | 0,947 |
| 121 | 0 | 0,093 | 0,366 | 179 | 0 | 0,068 | 0,277 |
| 97 | 1 | 0,154 | 0,542 | 93 | 0 | 0,529 | 0,919 |
| 118 | 0 | 0,056 | 0,227 | 75 | 1 | 0,378 | 0,840 |
| 111 | 1 | 0,945 | 0,997 | 101 | 0 | 0,050 | 0,203 |
| 21 | 1 | 0,093 | 0,367 | 169 | 0 | 0,050 | 0,204 |
| 162 | 0 | 0,010 | 0,034 | 12 | 0 | 0,134 | 0,491 |

TABEAU II.2 : PREDICTIONS DE LA PROBABILITE DE REUSSITE PAR LA TRANSFORMATION LOGISTIQUE ET LA TRANSFORMATION DE WARNER DE LA FONCTION DE PROBABILITE LINEAIRE.

Si nous calculons pour chaque modèle, la distance entre les valeurs observées et les valeurs prédites, nous avons :

prédictions transformation logistique : 6,15
 prédictions transformation de Warner : 8,65

Ainsi, les prédictions obtenues par la transformation logistique sont meilleures que celles résultant de la transformation de Warner de la fonction de probabilité linéaire. Cependant, dans chacun des cas, la distance valeurs observées-valeurs prédites est supérieure à celle obtenue lors de l'estimation du modèle linéaire par les moindres carrés ordinaires ou par les moindres carrés généralisés.

CHAPITRE III

LE MODELE LOGISTIQUE A VARIABLE ENDOGENE DICHOTOMIQUE

Pour éviter les inconvénients des méthodes présentées précédemment, il est nécessaire de postuler ab origine une fonction non décroissante et d'estimer les paramètres inconnus de cette fonction par les méthodes classiques d'inférence statistique, sans transformation du modèle, ni utilisation de données groupées artificiellement. Par sa simplicité, eu égard aux nombreuses formulations possibles, le modèle logistique s'impose d'emblée. Son estimation par la méthode du maximum de vraisemblance permet d'obtenir des estimateurs asymptotiquement efficaces et de construire des tests de significativité "exacts" pour les variables supposées déterminer le phénomène étudié.

III.1. CONSTRUCTION DE LA FONCTION DE VRAISEMBLANCE

Il est possible de distinguer deux catégories de statisticiens, les "anciens" et les "classiques". Les premiers admettent l'idée d'une inférence statistique reposant sur des connaissances et des données a priori. Les seconds, quant à eux, consentent un a priori seulement dans la forme analytique des lois de probabilités et rejettent tout a priori dans les méthodes d'inférence.

Le principe de la méthode du maximum de vraisemblance, procédé d'estimation développé par le courant classique de la statistique et utilisé pour l'estimation du modèle logistique, est le suivant. La fonction de densité de probabilité jointe de la variable aléatoire y , considérée comme fonction des paramètres inconnus $\beta' = (\beta_0, \dots, \beta_K)$ est appelée fonction de vraisemblance. Soit $L(y, \beta')$ cette fonction. Fisher a proposé, lorsque l'on dispose de l'observation y , d'estimer la valeur β_K inconnue par la valeur $\hat{\beta}_K(y)$ du paramètre maximisant la vraisemblance de l'échantillon, c'est-à-dire que :

$$L[y, \hat{\beta}_K(y)] \geq L[y, \beta_K] \quad \forall \beta_K$$

Le problème à résoudre désormais est celui de la spécification de la fonction de densité de probabilité jointe des observations y_i . Rappelons que nous estimons le modèle :

$$y_i = \frac{1}{1 + e^{-x_i \beta}} + \epsilon_i$$

Les hypothèses relatives aux erreurs sont les mêmes que celles décrites au chapitre I, à savoir :

$$E(\epsilon_i) = 0 \quad \forall i$$

$$E(\epsilon_i \epsilon_j) = y_i(1 - y_i) \quad \text{pour } i = j$$

$$= 0 \quad \text{sinon.}$$

Sous l'hypothèse de nullité de l'espérance mathématique de l'erreur aléatoire, la probabilité de réalisation de l'évènement expliqué est égale à :

$$\text{Prob}\{y_i=1\} = E(y_i | x_i; \beta) = \frac{1}{1+e^{-x_i\beta}}$$

L'erreur aléatoire ϵ_i ne peut prendre que deux valeurs, à savoir :

$$\epsilon_i = \frac{e^{-x_i\beta}}{1+e^{-x_i\beta}} \quad \text{si } y_i = 1$$

$$\epsilon_i = -\frac{1}{1+e^{-x_i\beta}} \quad \text{si } y_i = 0$$

Sachant que :

$$E(\epsilon_i) = \text{Prob}\left\{\epsilon_i = \frac{e^{-x_i\beta}}{1+e^{-x_i\beta}}\right\} \left(\frac{e^{-x_i\beta}}{1+e^{-x_i\beta}}\right) + \text{Prob}\left\{\epsilon_i = -\frac{1}{1+e^{-x_i\beta}}\right\} \left(-\frac{1}{1+e^{-x_i\beta}}\right) = 0$$

et que :

$$\text{Prob}\left\{\epsilon_i = \frac{e^{-x_i\beta}}{1+e^{-x_i\beta}}\right\} + \text{Prob}\left\{\epsilon_i = -\frac{1}{1+e^{-x_i\beta}}\right\} = 1$$

par un calcul analogue à celui présenté page 8, nous obtenons alors les expressions suivantes pour les probabilités de réalisation de l'erreur aléatoire ϵ_i :

$$\text{Prob}\left\{\epsilon_i = \frac{e^{-x_i\beta}}{1+e^{-x_i\beta}}\right\} = \frac{1}{1+e^{-x_i\beta}}$$

$$\text{et } \text{Prob}\left\{\epsilon_i = -\frac{1}{1+e^{-x_i\beta}}\right\} = \frac{e^{-x_i\beta}}{1+e^{-x_i\beta}}$$

Ainsi, il vient :

$$f(\epsilon_i) = \frac{1}{1+e^{-x_i\beta}} \quad \text{quand } y_i = 1$$

$$\text{et } f(\epsilon_i) = \frac{e^{-x_i\beta}}{1+e^{-x_i\beta}} \quad \text{quand } y_i = 0$$

De ce fait nous déduisons

$$f(y_i) = \frac{1}{1+e^{-x^i\beta}} \quad \text{pour } y_i = 1$$

et

$$f(y_i) = \frac{e^{-x^i\beta}}{1+e^{-x^i\beta}} \quad \text{pour } y_i = 0$$

Chaque variable aléatoire y_i est une variable aléatoire binomiale indicatrice. La fonction de densité de probabilité d'une telle variable peut s'écrire ainsi :

$$f(y_i) = \left(\frac{1}{1+e^{-x^i\beta}} \right)^{y_i} \left(\frac{e^{-x^i\beta}}{1+e^{-x^i\beta}} \right)^{1-y_i}$$

y_i étant indépendant de y_j , la fonction de densité de probabilité jointe de y n'est rien d'autre que le produit des fonctions de densité de probabilité individuelle. La fonction de vraisemblance de l'échantillon s'écrit alors de la manière suivante :

$$L(\beta_0, \dots, \beta_k | y_1, \dots, y_i, \dots, y_n) = \prod_{i=1}^n \left(\frac{1}{1+e^{-x^i\beta}} \right)^{y_i} \left(\frac{e^{-x^i\beta}}{1+e^{-x^i\beta}} \right)^{1-y_i}$$

III.2. ESTIMATION ET TESTS D'HYPOTHESES

L'estimation du modèle par la méthode du maximum de vraisemblance revient à maximiser la fonction $L(\beta_0, \dots, \beta_k | y_1, \dots, y_i, \dots, y_n)$ par rapport à tous les paramètres inconnus β_k . La condition pour avoir un maximum est que les dérivées premières de la vraisemblance par rapport aux paramètres inconnus soient nulles.

Habituellement, dans le cas linéaire, la résolution du système linéaire d'équations normales permet de déduire ces estimateurs. Il est bien évident que dans le cas qui nous préoccupe, ces équations ne sont pas linéaires dans les paramètres, de ce fait la résolution du système d'équations normales n'est pas simple. Seule une méthode d'optimisation numérique permet alors de découvrir les

estimateurs des paramètres inconnus. La fonction à maximiser, $L(\beta_0, \dots, \beta_k | y_1, \dots, y_1, \dots, \beta_k)$, étant convexe, nous sommes certains de trouver un maximum global si bien que les estimateurs des paramètres inconnus du modèle possèdent toutes les caractéristiques des estimateurs du maximum de vraisemblance. Ainsi ces estimateurs sont convergents, c'est-à-dire que :

$$\text{plim} \hat{\beta}_k = \beta_k$$

Leur variance asymptotique se définit de la façon suivante :

$$\text{Var.Asympt.}(\hat{\beta}_k) = \frac{1}{n} \lim_{n \rightarrow \infty} n \text{Var}(\hat{\beta}_k)$$

Les estimateurs du maximum de vraisemblance sont asymptotiquement efficients. Ainsi, tout autre estimateur convergent β_k^* de β_k a une variance asymptotique supérieure à celle de $\hat{\beta}_k$.

Outre ces propriétés, ces estimateurs sont asymptotiquement normaux. Ayant découvert par une méthode d'optimisation l'estimation des paramètres inconnus du modèle, nous pouvons pour juger de la significativité d'une variable x_k procéder de deux manières différentes. La première consiste à calculer le rapport de l'estimateur β_k à sa variance asymptotique définie par :

$$\text{Var.Asympt.}(\hat{\beta}_k) = - \left[\frac{\partial^2 L}{\partial \beta_k^2} / \beta_k = \hat{\beta}_k \right]^{-1}$$

et à comparer ce rapport à un t de Student.

La seconde consiste à utiliser le test du rapport de vraisemblance, à savoir :

$$\lambda = \frac{L_e}{L}$$

où L représente la valeur de la fonction de vraisemblance au point $\hat{\beta}$ et L_e représente la valeur de la fonction de vraisemblance au point

$$\hat{\beta}'_e = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{k-1}, 0)$$

Or

$$-2 \ln \lambda \underset{\text{Asympt.}}{\sim} \chi^2_{(1)}$$

La comparaison de la quantité $-2 \ln \lambda$ avec un χ^2 théorique permet alors de déterminer la significativité de la variable en question.

III.3. METHODES D'OPTIMISATION

Supposons une fonction $f(x_1, \dots, x_n)$ continue et dérivable. Les conditions suffisantes pour que $f(x_1, \dots, x_n)$ admette un maximum sont les suivantes :

- . les dérivées premières $\frac{\partial f}{\partial x_i}$ sont nulles,
- . le Hessien $\left\| \frac{\partial^2 f}{\partial x_i \partial x_j} \right\|$ est une matrice définie négative.

En pratique il est très difficile de résoudre les équations normales de sorte qu'il est nécessaire de recourir à des processus numériques d'optimisation. La majeure partie de ceux-ci consiste à choisir un point de départ et à procéder par itération selon le schéma ci-dessous, jusqu'à ce qu'un certain critère de convergence soit atteint:

$$x^{P+1} = x^P + h^P D^P$$

où x^P est l'approximation du maximum à la $P^{\text{ième}}$ itération.

D^P est un vecteur direction

h^P est un scalaire positif.

Parmi les nombreuses méthodes du gradient, nous exposons ci-après, peut-être celle qui est la plus connue, à savoir la méthode de Newton.

Appelons F_a le gradient $\left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)$ évalué au point (a_1, \dots, a_n) et S_a la matrice des dérivées secondes partielles évaluées en ce même point. Soit $x^0 = (x_1^0, \dots, x_n^0)$ un point de départ. Iterons selon le schéma

$$x^{P+1} = x^P + F(x^P)$$

Supposons que $f(x)$ admette un développement de Taylor au voisinage de x^P . Nous avons alors :

$$f(x^{P+1}) = f(x^P) + (x^{P+1} - x^P)' F(x_P) + \frac{1}{2} (x^{P+1} - x^P)' S_{xP} (x^{P+1} - x^P)$$

En maximisant cette approximation quadratique par rapport au point inconnu x^{P+1} , nous obtenons :

$$F(x^{P+1}) = F(x^P) + S_{xP} (x^{P+1} - x^P) = 0$$

D'où nous pouvons tirer la valeur de x^{P+1} . En effet, nous avons :

$$F(x^P) + S_{xP} (x^{P+1} - x^P) = 0$$

En multipliant à gauche par S_{xP}^{-1} , nous obtenons, en supposant que l'inverse de la matrice existe :

$$S_{xP}^{-1} F_{xP} + (x^{P+1} - x^P) = 0$$

$$D'où $x^{P+1} = x^P - S_{xP}^{-1} F_{xP}$$$

Cette expression constitue le processus itératif de la méthode de Newton.

En résumé, étant donné une fonction non linéaire $f(x)$, nous calculons F et S . Soit x^P un point de départ. Nous évaluons alors F_{xP} et S_{xP} . Si S_{xP} est définie négative, alors x^P est le maximum de la fonction, sinon nous calculons x^{P+1} et ainsi de suite.

Les méthodes de gradient nécessitent le calcul des dérivées premières et secondes de la fonction à maximiser. Certaines méthodes, comme celle des variations locales n'ont pas recours aux calculs des dérivées, ce qui présente un avantage certain lorsque la fonction à maximiser est complexe. Le principe de cette méthode est le suivant. Supposons que nous cherchions les valeurs x_1^* et x_2^* qui maximisent la fonction $f(x_1, x_2)$. Pour ce faire, nous nous donnons un point de départ (x_1^0, x_2^0) auquel est associée la valeur $f_1 = f(x_1^0, x_2^0)$ de la

fonction. Soit α la perturbation ou le pas initial que l'on accepte sur l'une quelconque des deux variables. Imaginons que nous fassions tout d'abord varier x_1 de $\pm \alpha$. Il est possible de calculer

$$\int_1^+ = f(x_1^0 + \alpha, x_2^0) \text{ et } \int_1^- = f(x_1^0 - \alpha, x_2^0)$$

La méthode des variations locales consiste alors à retenir pour nouvelle valeur de la variable x_1 , celle qui réalise le maximum de $\left\{ \int_1^-, \int_1^+ \right\}$. Soit x_1^α cette valeur. Il suffit ensuite de remplacer x_1^0 par x_1^α et d'itérer en acceptant cette fois-ci une perturbation sur la variable x_2 . Dès que nous trouvons un point stationnaire, c'est-à-dire un point tel qu'il n'est plus possible d'augmenter la valeur de la fonction dans une quelconque direction grâce au pas initial α , nous recommençons le processus en divisant la perturbation par deux. L'optimum est atteint lorsque la différence entre les valeurs de la fonction pour deux points stationnaires consécutifs est inférieure à un seuil donné.

III.4. EXEMPLE D'APPLICATION

Afin de comparer les résultats obtenus par les diverses méthodes d'estimation, nous reprenons ici l'estimation de la probabilité de réussite des étudiants-médecins à partir de l'échantillon présenté au chapitre I. L'optimisation de la fonction de vraisemblance du modèle logistique a été effectuée au moyen de la méthode des variations locales¹. Les résultats obtenus sont les suivants :

¹ L'optimisation de la fonction a nécessité 922 itérations, soit 1 h.15 d'utilisation de l'ordinateur PDP 15.

| VARIABLE | COEFFICIENT |
|--------------------------|-------------------------|
| Taille de la commune | -0,317*** |
| Revenus des parents/1000 | 0,117*** |
| Age/10 | -3,609*** |
| Test logique | 0,011* |
| Test de personnalité/10 | -0,251** |
| Moyenne à l'écrit du bac | 0,535*** |
| Etudes précédentes | 1,851*** |
| Origine du secondaire | 0,368* |
| Baccalauréat C | 1,412*** |
| Baccalauréat A,B,F,G | -15,100* |
| Constante | -0,496 |
| | % de variance expliquée |
| | 0,350 |

TABLEAU 3.1.: ESTIMATION DE LA FONCTION DE PROBABILITE LOGISTIQUE
PAR LA METHODE DU MAXIMUM DE VRAISEMBLANCE.

Les tests de significativité des variables sont basés sur le rapport de vraisemblance. Les seuils retenus sont les suivants:

* = 10 % ** = 5 % *** = 1 %

L'explication de la réussite des étudiants par le modèle logistique est supérieure de 5 ou de 7 % à celle obtenue par le modèle linéaire estimé par les moindres carrés ordinaires ou par les moindres carrés généralisés (cf. tableaux I.1. et I.2., chapitre I). Si ce point est important, un autre l'est encore plus pour le chercheur empiriste, il s'agit du problème de la significativité des variables. Alors qu'au vu de l'estimation de la fonction de probabilité linéaire, nous sommes amenés à rejeter l'influence de certaines variables sur la réussite, il n'en est plus de même dans le cas du modèle logistique. La raison en est qu'il était abusif d'admettre que les estimateurs des paramètres inconnus suivaient une loi de Student.

Non seulement certaines variables ne sont pas significatives à l'issue de l'estimation de la fonction de probabilité linéaire,

mais de plus la structure ordinale des variables influant sur la probabilité de réussite diffère considérablement selon que l'on adopte le modèle linéaire ou le modèle logistique.

Pour nous permettre d'évaluer les différences entre les prédictions obtenues par le modèle linéaire et par le modèle logistique, nous donnons ci-dessous les valeurs calculées de la probabilité de réussite pour chacun des individus constituant le sous-échantillon aléatoire défini précédemment. Les valeurs prédites sont obtenues à partir du tableau I.2. pour la fonction de probabilité linéaire et du tableau III.1. pour la fonction de probabilité logistique.

| N° observation | Valeur observée | Prédiction fonction linéaire | Prédiction fonction logistique | N° observation | Valeur observée | Prédiction fonction linéaire | Prédiction fonction logistique |
|----------------|-----------------|------------------------------|--------------------------------|----------------|-----------------|------------------------------|--------------------------------|
| 159 | 1 | 0,665 | 0,869 | 67 | 0 | 0,242 | 0,146 |
| 189 | 0 | -0,331 | 0,021 | 146 | 0 | -0,099 | 0,000 |
| 76 | 0 | -0,014 | 0,030 | 163 | 1 | 0,563 | 0,639 |
| 167 | 1 | 0,449 | 0,474 | 47 | 1 | 0,300 | 0,210 |
| 7 | 0 | -0,036 | 0,000 | 106 | 0 | 0,289 | 0,196 |
| 34 | 1 | 0,941 | 0,965 | 102 | 0 | 0,447 | 0,414 |
| 81 | 0 | 0,461 | 0,439 | 197 | 1 | 0,614 | 0,728 |
| 20 | 1 | 0,647 | 0,782 | 33 | 0 | 0,198 | 0,111 |
| 66 | 1 | 0,646 | 0,728 | 29 | 1 | 0,452 | 0,352 |
| 187 | 0 | 0,278 | 0,192 | 178 | 0 | 0,608 | 0,699 |
| 5 | 0 | -0,025 | 0,021 | 2 | 0 | 0,082 | 0,055 |
| 212 | 0 | 0,200 | 0,133 | 158 | 0 | 0,189 | 0,105 |
| 44 | 0 | 0,359 | 0,222 | 103 | 0 | 0,573 | 0,661 |
| 184 | 1 | 0,449 | 0,443 | 91 | 1 | 0,592 | 0,670 |
| 121 | 0 | 0,094 | 0,046 | 179 | 0 | 0,035 | 0,035 |
| 97 | 1 | 0,198 | 0,109 | 93 | 0 | 0,525 | 0,567 |
| 118 | 0 | -0,002 | 0,026 | 75 | 1 | 0,414 | 0,368 |
| 111 | 1 | 1,017 | 0,978 | 101 | 0 | -0,023 | 0,026 |
| 21 | 1 | 0,095 | 0,060 | 169 | 0 | -0,023 | 0,021 |
| 162 | 0 | -0,312 | 0,002 | 12 | 0 | 0,168 | 0,142 |

TABLEAU III.2.: PREDICTIONS DE LA PROBABILITE DE REUSSITE PAR LA FONCTION DE PROBABILITE LINEAIRE ET LA FONCTION DE PROBABILITE LOGISTIQUE.

La distance entre les valeurs observées et les valeurs prédites pour le modèle logistique est égale à 5,99 alors qu'elle est de 6,06 pour le modèle linéaire estimé par les moindres carrés généralisés. La comparaison des prédictions indique que par rapport au modèle logistique, le modèle linéaire surestime la probabilité de réussite des individus dont la variable endogène est égale à un, dans 50 % des cas, alors qu'il surestime la probabilité d'échec des individus dont la variable endogène est égale à zéro, dans 48 % des cas.

La comparaison des prédictions obtenues par la méthode de transformation logistique de la fonction de probabilité linéaire (tableau II.2., chapitre II) et par le modèle logistique indique quant à elle que par rapport à ce dernier, la transformation logistique de la fonction de probabilité linéaire surestime la probabilité d'échec et sous-estime la probabilité de réussite.

La distance prédictions-observations résultant de la méthode de transformation de Warner de la fonction de probabilité linéaire est 1,44 fois plus élevée que la distance prédictions-observations issue du modèle logistique.

Le tableau ci-dessous donne l'élasticité de la probabilité de réussite (calculée au point moyen) par rapport à chacune des variables dans le but de faciliter la comparaison des résultats fournis par le modèle linéaire estimé par les moindres carrés généralisés et par le modèle logistique estimé par la méthode du maximum de vraisemblance. L'avantage qu'il y a à comparer les élasticités plutôt que les effets marginaux tient au fait que dans le modèle logistique ceux-ci ne sont pas constants comme dans le modèle linéaire, mais varie en fonction du niveau de probabilité auquel on se situe.

| VARIABLES | Modèle logistique | Modèle linéaire |
|--------------------------|-------------------|-----------------|
| Taille de la commune | - 0,54 | - 0,23 |
| Revenus des parents/1000 | 0,42 | 0,23 |
| Age/10 | - 4,88 | - 2,88 |
| Test logique | 0,23 | 0,10 |
| Test de personnalité/10 | - 0,84 | - 0,38 |
| Moyenne à l'écrit du bac | 4,15 | 3,00 |
| Etudes précédentes | 0,07 | 0,04 |
| Origine du secondaire | 0,22 | 0,22 |
| Baccalauréat C,D | 0,35 | 0,26 |
| Baccalauréat A,B,F,G,D | - 0,55 | - 0,02 |

TABLEAU III.3.: ELASTICITES DE LA PROBABILITE DE REUSSITE PAR RAPPORT A CHACUNE DES VARIABLES.

Les variables influant le plus sur la probabilité de réussite (du point de vue des élasticités) sont dans l'un et l'autre modèle les variables "Age" et "Moyenne à l'écrit du baccalauréat". Toutefois, les élasticités de la probabilité de réussite par rapport à ces variables sont beaucoup plus faibles dans le modèle linéaire comme le sont d'ailleurs toutes les autres élasticités. Alors qu'une augmentation identique de chacune des variables exogènes aurait pour effet de laisser pratiquement inchangée la probabilité de réussite du modèle linéaire, elle diminuerait de plus de 1 % la probabilité du modèle logistique.

CHAPITRE IV

LE MODELE LOGISTIQUE A VARIABLE ENDOGENE POLYTOMIQUE

Les chapitres précédents ont été consacrés à l'étude du modèle à variable dépendante dichotomique. Il s'avère alors que le modèle logistique, estimé par la méthode du maximum de vraisemblance, procure les meilleurs résultats tant au point de vue de la qualité des estimateurs des paramètres inconnus, qu'au point de vue des prédictions obtenues.

Naturellement, il y a beaucoup de situations dans lesquelles la variable dépendante d'un modèle est polytomique, c'est-à-dire qu'elle admet plus de deux modalités. Le but de ce chapitre est de présenter la généralisation du modèle logistique à variable dépendante dichotomique au modèle à variable polytomique.

IV.1. PRESENTATION DU MODELE

Imaginons un bachelier i décidé à poursuivre des études supérieures universitaires. Cet individu est alors placé devant un choix. En effet, il lui faut résoudre le problème du lieu de déroulement de ses études. Un certain nombre d'Universités s'offrent à lui et parmi celles-ci il doit en choisir une. Supposons qu'il existe seulement deux Universités, U_1 et U_2 . L'évènement

$E = \{\text{lieu de déroulement des études supérieures}\}$ est dans ce cas une variable dichotomique. Notons

$y_i = 1$, si l'individu i choisit l'Université U_1

$y_i = 0$, si l'individu i choisit l'Université U_2

Supposons que la variable y_i soit déterminée par k variables exogènes indépendantes et fixes, x (binaires ou non). Notons y_i le choix de l'individu i , $i=1, \dots, N$. En reprenant la notation précédemment utilisée, nous exprimons la probabilité de se rendre à l'Université U_1 , de la manière suivante :

$$\text{Prob}\{y_i = U_1\} = \frac{1}{1 + e^{-x_i \beta}} \quad i = 1, \dots, N$$

dans lequel β est le vecteur d'ordre $(k+1, 1)$ des paramètres inconnus, à savoir $\beta' = (\beta_0, \beta_1, \dots, \beta_k)$. Plutôt que d'estimer ce modèle, il eût été possible d'estimer le modèle

$$\text{Prob}\{y_i = U_2\} = \frac{1}{1 + e^{-x_i \beta^*}} \quad i = 1, \dots, N$$

dans lequel β est le vecteur d'ordre $(k+1, 1)$ des paramètres inconnus à savoir $\beta' = (\beta_0^*, \beta_1^*, \dots, \beta_k^*)$ en posant cette fois-ci :

$y_i = 1$ si l'individu i choisit l'Université U_2 .

$y_i = 0$ si l'individu i choisit l'Université U_1

Naturellement, les valeurs absolues des paramètres inconnus de ce modèle sont les mêmes que celles des paramètres inconnus du modèle initial, seuls les signes sont inversés. En effet, si la

la variation d'une variable exogène augmente la probabilité de se rendre à l'Université U_1 , elle diminue dans les mêmes proportions la probabilité de se rendre à l'Université U_2 , si bien que :

$$\beta_k + \beta_k^* = 0 \quad \forall k$$

puisque en fait

$$\text{Prob}\{y_i=U_2\} = 1 - \text{Prob}\{y_i=U_1\}$$

Il est possible d'écrire le modèle

$$\text{Prob}\{y_i=U_1\} = \frac{1}{1+e^{-x_i\beta}}$$

de la manière suivante :

$$\text{Prob}\{y_i=U_1\} = \frac{e^{x_i\beta}}{e^{x_i\beta} + e^{x_i\beta^*}}$$

Etant donné que $\beta = \beta^*$, nous avons

$$\text{Prob}\{y_i=U_1\} = \frac{1}{1 + e^{-2x_i\beta}}$$

La fonction de distribution logistique univariée s'écrivant

$$f(t_1) = \frac{1}{1+e^{-t_1}} \quad -\infty < t_1 < +\infty$$

aussi les probabilités $\text{Prob}\{y_i=U_1\}$ et $\text{Prob}\{y_i=U_2\}$ s'obtiennent en posant $t_1 = 2x_i\beta$

Imaginons à présent le cas plus réaliste où l'individu i doit choisir parmi un ensemble Q d'Universités, U_1, \dots, U_Q . Désormais l'évènement

$E = \{\text{lieu de déroulement des études supérieures}\}$

est une variable polytomique comportant autant de modalités qu'il existe d'Universités réparties sur le territoire national.

Soit alors

$$P_{ij} = \text{Prob}\{y_i = U_j\}$$

la probabilité que l'individu i choisisse comme lieu de déroulement de sa scolarité l'Université j . Par analogie avec le cas dichotomique précédemment évoqué nous avons donc

$$\sum_{j=1}^Q P_{ij} = 1 \quad \forall i, i=1, \dots, N$$

Utilisons la fonction logistique standardisée pour exprimer le lien entre le choix d'un individu i et le vecteur x^i des variables exogènes supposées déterminer sa préférence. Dans le cas multivariée, la fonction de distribution logistique s'écrit de la manière suivante :

$$f(t_1, \dots, t_m) = \frac{1}{1 + \sum_{j=1}^m e^{-t_j}} \quad -\infty < t_j < +\infty$$

Ainsi, en faisant l'analogie avec le modèle dichotomique, il est possible d'écrire la probabilité de la façon suivante :

$$P_{ij} = \frac{e^{x^i \beta_j}}{\sum_{k=1}^Q e^{x^i \beta_k}} \quad j = 1, \dots, Q ; i = 1, \dots, N$$

avec

$$\sum_{j=1}^Q P_{ij} = 1$$

et

$$\sum_{j=1}^Q \beta_j = 0$$

où x^i est le vecteur d'ordre $(1, k+1)$ des variables explicatives associées à l'individu i , β_j le vecteur d'ordre $(k+1, 1)$ des paramètres inconnus associés à la modalité j de l'évènement considéré.

IV.2. ESTIMATION DU MODELE

A l'instar du modèle à variable dépendante dichotomique, le modèle à variable endogène polytomique est lui aussi estimé par la méthode du maximum de vraisemblance. Alors que dans le premier cas nous devons estimer qu'une seule fonction de probabilité et donc qu'un seul vecteur de paramètres inconnus, dans le second il est nécessaire d'estimer autant de fonctions de probabilité que la variable polytomique peut prendre de modalités et par conséquent il faut découvrir $(Q \times k)$ paramètres inconnus.

Reprenons le cas dichotomique précédent dans lequel nous avons

$$\text{Prob}\{y_i=U_1\} = \frac{1}{1+e^{-x_i\beta}} \quad \text{et} \quad \text{Prob}\{y_i=U_2\} = \frac{e^{-x_i\beta}}{1+e^{-x_i\beta}}$$

La fonction de vraisemblance de ce modèle s'écrit :

$$L(\beta_0, \dots, \beta_k) = \prod_{i=1}^n \left(\frac{1}{1+e^{-x_i\beta}} \right)^{y_i} \left(\frac{e^{-x_i\beta}}{1+e^{-x_i\beta}} \right)^{1-y_i}$$

En appelant P_{i1} la probabilité que l'individu i se rende à l'Université U_1 et P_{i2} la probabilité qu'il se rende à l'Université U_2 , il vient alors

$$L(\beta_0, \dots, \beta_k) = \prod_{i=1}^n P_{i1}^{y_i} P_{i2}^{1-y_i}$$

Si l'individu i se rend à l'Université U_1 , notons

$$V_{i1} = 1 \quad \text{et} \quad V_{i2} = 0$$

Si au contraire l'individu i se rend à l'université U_2 , notons

$$V_{i1} = 0 \quad \text{et} \quad V_{i2} = 1$$

Nous pouvons alors écrire la fonction de vraisemblance du modèle de la manière suivante

$$L(\beta_0, \dots, \beta_k) = \prod_{i=1}^n P_{i1}^{V_{i1}} P_{i2}^{V_{i2}}$$

Soit encore de façon plus dense

$$L(\beta_0, \dots, \beta_k) = \prod_{i=1}^n \prod_{j=1}^2 P_{ij}^{V_{ij}}$$

L'estimation du modèle par la méthode du maximum de vraisemblance revient à maximiser la fonction $L(\beta_0, \dots, \beta_k)$ par rapport à tous les paramètres inconnus du modèle. Toutefois pour que tous ces $(Q \times k)$ paramètres soient définis il est nécessaire de maximiser cette fonction sous la contrainte

$$\prod_{j=1}^Q \beta_j = 0$$

où β_j est le vecteur des paramètres inconnus des variables exogènes associées à la $j^{\text{ième}}$ modalité du phénomène expliqué. La condition pour avoir un maximum est que les dérivées premières de la vraisemblance par rapport aux paramètres inconnus soient nulles. Comme dans le cas dichotomique, les équations normales n'étant pas linéaires, il est nécessaire de recourir à un processus d'optimisation. Toutefois, la fonction étant convexe, nous sommes certains de trouver un maximum global si bien que les estimateurs des paramètres inconnus du modèle possèdent toutes les caractéristiques des estimateurs du maximum de vraisemblance.

Les tests d'hypothèse sont construits de la même façon que dans le cas dichotomique c'est-à-dire soit en comparant le rapport du maximum de vraisemblance à un χ^2 théorique à un degré de liberté, soit en comparant le rapport de l'estimateur du paramètre inconnu à sa variance asymptotique à un t de Student.

Dans un modèle linéaire, le coefficient de détermination multiple donne une mesure de la plus ou moins bonne liaison (linéaire) qui existe entre les variables

explicatives dans leur ensemble et la variable expliquée. Or dans un modèle non linéaire, comme le modèle logistique, nous ne pouvons plus calculer un tel coefficient. Il est alors nécessaire d'effectuer un test de significativité sur l'ensemble des variables exogènes permettant ainsi de juger si les variables retenues déterminent de façon significative la probabilité de réalisation de l'évènement expliqué.

CHAPITRE V

LE MODELE LOGISTIQUE D'EQUATIONS SIMULTANÉES A VARIABLES ENDOGENES QUALITATIVES

L'introduction de variables endogènes qualitatives dans un modèle d'équations simultanées pose exactement les mêmes problèmes que ceux rencontrés lors de l'estimation du modèle simple à variable expliquée qualitative. Là encore, il est nécessaire de substituer à la forme linéaire généralement postulée une forme non linéaire, telle que la forme logistique, dont les paramètres inconnus sont estimés par la méthode du maximum de vraisemblance à information complète.

Pour faciliter l'exposé, nous considérons ci-après un modèle comportant seulement deux variables endogènes qualitatives. Après l'examen détaillé du modèle simultané à variables dépendantes dichotomiques nous généralisons ensuite au modèle à variables dépendantes polytomiques.

V.1. PRESENTATION DU MODELE

Un modèle économétrique est une formalisation d'un modèle économique exprimant, par un ensemble d'équations, certaines variables économiques en fonction d'elles-mêmes d'une part, en fonction d'autres variables économiques ou non d'autre part. La caractéristique d'un système simultané réside dans le fait qu'une variable expliquée d'une équation apparaît comme variable explicative dans une autre équation. Ces équations peuvent être de deux types, stochastiques ou comptables. En général, il est toujours possible d'éliminer ces dernières par substitution.

A titre d'illustration, considérons l'évènement

$$E = \{\text{emploi occupé par l'individu } i\}$$

Supposons que nous retenions deux types d'emploi, les emplois techniques d'une part, les emplois administratifs d'autre part. Nous définissons alors une variable dichotomique que nous codons :

$$y_{i1} = 1 \text{ si l'individu } i \text{ occupe un emploi technique}$$
$$y_{i1} = 0 \text{ si l'individu } i \text{ occupe un emploi administratif.}$$

Supposons que cette variable soit déterminée par k variables exogènes x , indépendantes et fixes et par une variable dichotomique y_2 , symbolisant l'évènement.

$E' = \{\text{secteur d'activité de l'individu } i\}$
codée

$$y_{i2} = 1 \text{ si le secteur d'activité de l'individu } i \text{ est le secteur de la production,}$$
$$y_{i2} = 0 \text{ si le secteur d'activité de l'individu } i \text{ est le secteur des services.}$$

Nous avons alors le modèle suivant :

$$y_{i1} = f(x^i, y_{i2}, \epsilon_{i1})$$

Sous l'hypothèse d'une représentation logistique du phénomène, il vient :

$$y_{i1} = \frac{1}{1 + e^{-x^i \beta - \alpha y_{i2}}} + \epsilon_{i1} \quad i = 1, \dots, n$$

où x^i est le vecteur d'ordre $(1, k+1)$ des variables exogènes déterminent la variable dichotomique y_{i1} , β le vecteur d'ordre $(k+1, 1)$ des paramètres inconnus de ces variables, y_{i2} la variable dichotomique représentant le secteur d'activité de l'individu i , α le coefficient de cette variable et ϵ_{i1} un terme d'erreur aléatoire.

Sous l'hypothèse

$$E(\epsilon_{i1}) = 0 \quad \forall i$$

nous avons alors

$$E(y_{i1} | x^i, y_{i2}) = \text{Prob}\{y_{i1} = 1 | x^i, y_{i2}\} = \frac{1}{1 + e^{-x^i \beta - \alpha y_{i2}}}$$

Imaginons à présent que nous cherchions à expliquer la variable dichotomique y_2 . Celle-ci est déterminée par les variables exogènes z mais également par la variable endogène y_1 . En effet, la probabilité de travailler dans un secteur d'activité plutôt que dans un autre est partiellement conditionnée par la probabilité d'exercer tel ou tel emploi. Aussi, toujours en acceptant l'hypothèse d'une représentation logistique, nous avons :

$$y_{i2} = \frac{1}{1 + e^{-z^i \gamma - S y_{i1}}} + \epsilon_{i2}$$

où Z^i est le vecteur d'ordre $(1, l+1)$ des variables exogènes déterminant la variable dichotomique y_{i2} , γ le vecteur d'ordre $(l+1, 1)$ des paramètres inconnus de ces variables, y_{i1} la variable dichotomique représentant l'emploi occupé par l'individu i , S le coefficient mesurant l'effet de cette variable sur le secteur d'activité et ϵ_{i2} un terme d'erreur aléatoire.

Sous l'hypothèse $E(\epsilon_{i2}) = 0 \quad \forall i$

et en considérant momentanément y_{i1} comme une variable exogène, nous avons alors :

$$E(y_{i2} | z^i, y_{i1}) = \text{Prob}\{y_{i2}=1 | z^i, y_{i1}\} = \frac{1}{1+e^{-z^i\gamma - \beta y_{i1}}}$$

En raison de la simultanéité du problème, il est impossible d'estimer les paramètres inconnus à partir de chaque équation considérée indépendamment l'une de l'autre. Ces estimations doivent être effectuées à partir de l'ensemble des équations, c'est-à-dire qu'il nous faut considérer le modèle simultané suivant :

$$\left. \begin{aligned} \text{Prob}\{y_{i1}=1 | y_{i2}\} &= \frac{1}{1+e^{-x_{i1}\beta - \alpha y_{i2}}} \\ \text{Prob}\{y_{i2}=1 | y_{i1}\} &= \frac{1}{1+e^{-z^i\gamma - \beta y_{i1}}} \end{aligned} \right\}$$

Ce système présente la particularité d'être à variables endogènes mutuellement dépendantes. Naturellement, il ne s'agit pas du modèle général d'équations simultanées. C'est dans un but purement didactique que nous avons choisi d'exposer ce modèle car sa généralisation à un système comportant plus de deux variables endogènes devient vite très complexe.

V.2. ESTIMATION DU MODELE

En raison du caractère mutuellement dépendant des variables endogènes il n'est pas nécessaire d'estimer $(k+1+2)$ paramètres mais seulement $(k+1+1)$ puisque les coefficients des variables endogènes dans chaque équation sont alors identiques. En effet, les probabilités conditionnelles des différents événements sont égales à :

$$\begin{aligned} \text{Prob}\{y_{i1}=1|y_{i2}=0\} &= \frac{\text{Prob}\{y_{i1}=1, y_{i2}=0\}}{\text{Prob}\{y_{i2}=0\}} = \frac{1}{1+e^{-x^i\beta}} \\ \text{Prob}\{y_{i1}=0|y_{i2}=0\} &= \frac{\text{Prob}\{y_{i1}=0, y_{i2}=0\}}{\text{Prob}\{y_{i2}=0\}} = \frac{e^{-x^i\beta}}{1+e^{-x^i\beta}} \\ \text{Prob}\{y_{i1}=1|y_{i2}=1\} &= \frac{\text{Prob}\{y_{i1}=1, y_{i2}=1\}}{\text{Prob}\{y_{i2}=1\}} = \frac{1}{1+e^{-x^i\beta-\alpha}} \\ \text{Prob}\{y_{i1}=0|y_{i2}=1\} &= \frac{\text{Prob}\{y_{i1}=0, y_{i2}=1\}}{\text{Prob}\{y_{i2}=1\}} = \frac{e^{-x^i\beta-\alpha}}{1+e^{-x^i\beta-\alpha}} \end{aligned}$$

D'où nous tirons

$$e^{-x^i\beta} = \frac{\text{Prob}\{y_{i1}=0, y_{i2}=0\}}{\text{Prob}\{y_{i1}=1, y_{i2}=0\}}$$

et

$$e^{-x^i\beta-\alpha} = \frac{\text{Prob}\{y_{i1}=0, y_{i2}=1\}}{\text{Prob}\{y_{i1}=1, y_{i2}=1\}}$$

En remplaçant dans cette dernière expression $e^{-x^i\beta}$ par le rapport des probabilités jointes correspondantes, nous obtenons :

$$e^{-\alpha} \frac{\text{Prob}\{y_{i1}=0, y_{i2}=0\}}{\text{Prob}\{y_{i1}=1, y_{i2}=0\}} = \frac{\text{Prob}\{y_{i1}=0, y_{i2}=1\}}{\text{Prob}\{y_{i1}=1, y_{i2}=1\}}$$

Soit

$$e^{\alpha} = \frac{\text{Prob}\{y_{i1}=1, y_{i2}=1\} \text{Prob}\{y_{i1}=0, y_{i2}=0\}}{\text{Prob}\{y_{i1}=0, y_{i2}=1\} \text{Prob}\{y_{i1}=1, y_{i2}=0\}}$$

De la même façon :

$$\text{Prob}\{y_{i2}=1|y_{i1}=0\} = \frac{\text{Prob}\{y_{i2}=1, y_{i1}=0\}}{\text{Prob}\{y_{i1}=0\}} = \frac{1}{1+e^{-z^i\gamma}}$$

$$\text{Prob}\{y_{i2}=0|y_{i1}=0\} = \frac{\text{Prob}\{y_{i2}=0, y_{i1}=0\}}{\text{Prob}\{y_{i1}=0\}} = \frac{e^{-z^i\gamma}}{1+e^{-z^i\gamma}}$$

$$\text{Prob}\{y_{i2}=1|y_{i1}=1\} = \frac{\text{Prob}\{y_{i2}=1, y_{i1}=1\}}{\text{Prob}\{y_{i1}=1\}} \frac{1}{1+e^{-z^1\gamma-S}}$$

$$\text{Prob}\{y_{i2}=0|y_{i1}=1\} = \frac{\text{Prob}\{y_{i2}=0, y_{i1}=1\}}{\text{Prob}\{y_{i1}=1\}} \frac{e^{-z^1\gamma-S}}{1+e^{-z^1\gamma-S}}$$

Par un calcul analogue au précédent, il vient :

$$e^S = \frac{\text{Prob}\{y_{i2}=1, y_{i1}=1\} \text{Prob}\{y_{i2}=0, y_{i1}=0\}}{\text{Prob}\{y_{i2}=0, y_{i1}=1\} \text{Prob}\{y_{i2}=1, y_{i1}=0\}}$$

Où

$$\alpha = S$$

Ainsi, le système d'équation simultanée peut se mettre sous la forme suivante :

$$\left. \begin{aligned} \text{Prob}\{y_{i1}=1|y_{i2}\} &= \frac{1}{1+e^{-x^1\beta-\alpha y_{i2}}} \\ \text{Prob}\{y_{i2}=1|y_{i1}\} &= \frac{1}{1+e^{-z^1\gamma-\alpha y_{i1}}} \end{aligned} \right\}$$

L'estimation de ce modèle d'équations simultanées par la méthode du maximum de vraisemblance à information complète nécessite au préalable la construction de la fonction de vraisemblance du système. Celle-ci n'est rien d'autre que le produit des probabilités jointes individuelles de réalisation des événements expliqués.

Ainsi la probabilité jointe des évènements $y_{i1}=0$ et $y_{i2}=0$ est égale à :

$$\text{Prob}\{y_{i1}=0, y_{i2}=0\} = \text{Prob}\{y_{i1}=0|y_{i2}=0\} \text{Prob}\{y_{i2}=0\}$$

$$\text{Or } \text{Prob}\{y_{i2}=0\} = \text{Prob}\{y_{i2}=0|y_{i1}=0\} + \text{Prob}\{y_{i2}=0|y_{i1}=1\} = \frac{1}{1+e^{z^1\gamma}} + \frac{1}{1+e^{z^1\gamma+\alpha}}$$

Comme

$$\text{Prob}\{y_{i1}=0|y_{i2}=0\} = \frac{1}{1+e^{x^1\beta}}$$

Nous avons donc

$$\text{Prob}\{y_{i1}=0, y_{i2}=0\} = \frac{1}{1+e^{x^{i1}\beta}} \left[\frac{1}{1+e^{z^{i1}\gamma}} + \frac{1}{1+e^{z^{i1}\gamma+\alpha}} \right]$$

En agissant de la même façon, nous obtenons :

$$\text{Prob}\{y_{i1}=1, y_{i2}=0\} = \frac{e^{x^{i1}\beta}}{1+e^{x^{i1}\beta}} \left[\frac{1}{1+e^{z^{i1}\gamma}} + \frac{1}{1+e^{z^{i1}\gamma+\alpha}} \right]$$

$$\text{Prob}\{y_{i1}=0, y_{i2}=1\} = \frac{1}{1+e^{x^{i1}\beta+\alpha}} \left[\frac{e^{z^{i1}\gamma}}{1+e^{z^{i1}\gamma}} + \frac{e^{z^{i1}\gamma+\alpha}}{1+e^{z^{i1}\gamma+\alpha}} \right]$$

$$\text{Prob}\{y_{i1}=1, y_{i2}=1\} = \frac{e^{x^{i1}\beta+\alpha}}{1+e^{x^{i1}\beta+\alpha}} \left[\frac{e^{z^{i1}\gamma}}{1+e^{z^{i1}\gamma}} + \frac{e^{z^{i1}\gamma+\alpha}}{1+e^{z^{i1}\gamma+\alpha}} \right]$$

Soit $\Theta_{m,n} = \{i | y_{i1}=m, y_{i2}=n\}$ $m, n = 0, 1$

La fonction de vraisemblance du système est alors égale à :

$$L(\beta, \gamma, \alpha) = \prod_{i=1}^n \prod_{m=0}^1 \prod_{n=0}^1 \text{Prob}\{y_{i1}=m, y_{i2}=n\}$$

L'estimation des paramètres inconnus du système par la méthode du maximum de vraisemblance revient à maximiser la fonction $L(\beta, \gamma, \alpha)$ par rapport à tous les paramètres inconnus du modèle. Pour ce faire, il est nécessaire de recourir, comme dans les cas précédemment étudiés, à un processus d'optimisation numérique. La façon la plus directe de mener les tests de significativité des variables est d'utiliser le rapport du maximum de vraisemblance.

V.3. GENERALISATION AU CAS POLYATOMIQUE.

Plutôt que d'envisager deux catégories d'emploi, imaginons-en m . De même, plutôt que de retenir deux secteurs d'activités supposons que nous en définissions n . Ainsi, les variables qualitatives y_{i1} et y_{i2} sont des variables polytomiques ayant respectivement m et n modalités. Dans ces conditions, le système d'équations simultanées

s'écrit :

$$\text{Prob}\{y_{i1}=r|y_{i2}=s\} = \frac{1}{1+e^{-x^i \beta_r - \sum_{s=1}^m \alpha_{sr} y_{i2}^s}} \quad r = 1, m$$

$$\text{Prob}\{y_{i2}=s|y_{i1}=r\} = \frac{1}{1+e^{-z^i \gamma_s - \sum_{r=1}^m \alpha_{rs} y_{i1}^r}} \quad s = 1, n$$

où x^i et z^i représentent respectivement les vecteurs d'ordre $(1, k+1)$ et $(1, l+1)$ des variables exogènes associées à l'individu i , β_r le vecteur d'ordre $(k+1, 1)$ des paramètres inconnus associés à la $r^{\text{ième}}$ modalité de la variable y_{i1} , α_{sr} l'influence de la $s^{\text{ième}}$ modalité de la variable y_{i2} sur la $r^{\text{ième}}$ modalité de la variable y_{i1} , γ_s le vecteur d'ordre $(l+1, 1)$ des paramètres inconnus associés à la $s^{\text{ième}}$ modalité de la variable y_{i2} et α_{rs} l'influence de la $r^{\text{ième}}$ modalité de la variable y_{i1} sur la $s^{\text{ième}}$ modalité de la variable y_{i2} .

Ainsi il est nécessaire d'estimer $(k \times m)$ paramètres β , $(l \times n)$ paramètres γ et $(m \times n)$ paramètres α en raison du caractère mutuellement dépendant des variables y_{i1} et y_{i2} .

Comme précédemment, il faut pour découvrir les estimateurs des paramètres inconnus du système par la méthode du maximum de vraisemblance à information complète, définir au préalable les probabilités individuelles jointes de réalisation des différents événements. Ainsi, d'après le théorème des probabilités conditionnelles, la probabilité de réalisation de l'événement joint $\{y_{i1}=r, y_{i2}=s\}$ est égale à :

$$\text{Prob}\{y_{i1}=r, y_{i2}=s\} = \text{Prob}\{y_{i1}=r|y_{i2}=s\} \text{Prob}\{y_{i2}=s\}$$

$$\text{Or } \text{Prob}\{y_{i1}=r|y_{i2}=s\} = \frac{1}{1+e^{-x^i \beta_r - \alpha_{sr}}}$$

D'autre part nous avons

$$\text{Prob}\{y_{i2}=s\} = \sum_{r=1}^m \text{Prob}\{y_{i2}=s | y_{i1}=r\}$$

Etant donné que

$$\text{Prob}\{y_{i2}=s | y_{i1}=r\} = \frac{1}{1 + e^{-z^i y_s - \alpha_{sr}}}$$

Alors

$$\text{Prob}\{y_{i2}=s\} = \sum_{r=1}^m \frac{1}{1 + e^{-z^i y_s - \alpha_{sr}}}$$

D'où

$$\text{Prob}\{y_{i1}=r, y_{i2}=s\} = \frac{e^{x^i \beta + \alpha_{sr}}}{1 + e^{x^i \beta_r + \alpha_{sr}}} \frac{\sum_{r=1}^m e^{z^i y_s + \alpha_{sr}}}{\sum_{r=1}^m e^{z^i y_s + \alpha_{sr}}}$$

Les probabilités de réalisation des autres évènements joints s'obtiennent de la même façon. Contrairement au système d'équations simultanées à variables endogènes dichotomiques, il ne saurait être question pour nous d'énumérer toutes ces probabilités, aussi construirons-nous la fonction de vraisemblance du système à partir de la probabilité de réalisation de l'évènement joint général ci-dessus.

Soit

$$O_{rs} = \{i | y_{i1}=r, y_{i2}=s\} \quad r = 1, \dots, m \text{ et } s = 1, \dots, n$$

La fonction de vraisemblance du système est alors égale à

$$L(\beta, \gamma, \alpha) = \prod_{i=1}^n \prod_{r=1}^m \prod_{s=1}^n \text{Prob}\{y_{i1}=r, y_{i2}=s\}$$

Par un processus d'optimisation numérique, il est alors possible de découvrir les estimateurs des paramètres inconnus du système et par là même de construire des tests de significativité des variables supposées déterminer le phénomène étudié.

CONCLUSION

Une fois encore, preuve est faite que la routine n'est pas du domaine de l'économétrie. En choisissant l'habituelle forme linéaire pour représenter la liaison entre une variable dépendante qualitative et un certain nombre de variables exogènes, le chercheur empiriste est certain d'obtenir des prédictions de probabilités biaisées et des tests de significativité des variables "inexactes". Etre conscient de cet état est d'importance, avoir la volonté d'y remédier l'est encore plus lorsqu'il s'agit de déduire des politiques à partir de considérations empiriques.

Plusieurs modèles sont alors envisageables pour représenter une probabilité et pour en donner une estimation. Sans conteste le modèle logistique est préférable à tout autre. En effet, les techniques d'estimation de données groupées, basées sur la formulation de modèles non linéaires ex post ou ex ante présentent l'inconvénient de réduire l'information d'une part, de manquer quelque peu de rigueur économétrique d'autre part, dans la mesure où les problèmes de transformation d'un modèle non linéaire en un modèle linéaire sont quasiment éludés. Le modèle logistique à variable dépendante qualitative permet quant à lui d'éviter ces deux inconvénients et présente l'avantage de fournir de meilleures prédictions de la probabilité de réalisation de l'évènement expliqué. Son estimation par la méthode du maximum de vraisemblance nécessite le recours à un processus d'optimisation numérique. Quelle que soit la nature de la fonction à optimiser, de tels processus sont toujours

difficiles à mettre en oeuvre. Selon la nature de cette fonction, les résultats qu'ils procurent sont plus ou moins fiables. Dans le cas du modèle logistique, la fonction de vraisemblance étant convexe, elle admet alors un maximum global, si bien que les estimateurs des paramètres inconnus possèdent toutes les propriétés des estimateurs du maximum de vraisemblance. Ils sont donc convergents, asymptotiquement efficaces et asymptotiquement normaux.

La généralisation du modèle logistique à variable dépendante dichotomique au modèle à variable dépendante polytomique permet quant à elle de traiter des situations plus variées qu'il serait évidemment possible de dichotomiser au prix toutefois d'une schématisation certaine de la réalité. L'estimation de ce modèle ne pose toutefois pas plus de problème que l'estimation du précédent, si ce n'est qu'il faut estimer un nombre plus important de paramètres inconnus pour un même nombre de variables exogènes.

Le système d'équations logistiques simultanées à variables endogènes qualitatives permet quant à lui de formaliser des situations très fréquentes en économie. Traiter séparément les équations qui le compose reviendrait à estimer des modèles simples à variables dépendantes dichotomiques ou polytomiques, et par là même à introduire un biais dans la valeur des estimateurs des paramètres inconnus, tout comme cela arrive avec le système d'équations simultanées à variables endogènes continues.

BIBLIOGRAPHIE

- AIGNER D.S., GOLDBERGER A.S. et KALTON G. "On the Explanatory Power of Dummy Variable Regressions" *International Economic Review*, 16, 1975, pp.503-510.
- ASHFORD J.R., SOWDEN R.R. "Multivariate Probit Analysis", *Biometrics*, 26, 1970, pp.535-546.
- ASHTON, W.D. *The Logit Transformation*, Hafner, New-York, 1972.
- GOLDBERGER, A.S. *Econometric Theory*, Wiley, New-York, 1972.
- GRIZZLE, J.E. "Multivariate Logit Analysis", *Biometrics*, 27, 1971, pp.1057-1062.
- GUNDERSON, M. *Determinants of Individual Success in on the Job Training : An Econometric Study*. Ph.D. Dissertation, University of Wisconsin, 1971.
- GUNDERSON, M. "Retention of Trainees. A study with Dichotomous Dependant Variables" *Journal of Econometrics*, 2, 1974, pp.79-93.
- HODGES, J.L. "Fitting the Logistic by Maximum Likelihood", *Biometrics*, 14, 1958, pp.453-461.
- JOHNSON, T. "Qualitative and Limited Dependent Variables in Economic Relationships", *Econometrica*, 40, 1972, pp.455-462.
- LASSIBILLE, G. *L'Analyse des Transformations en Econométrie*, Mémoire dactylographié, Université de Dijon, 1975.
- LASSIBILLE, G. "L'estimation de Modèle à Variable Dépendante Dichotomique", à paraître in la Revue *Economie Appliquée*. 2ème trimestre 1979.
- McGILLIVRAY, R.G. "Estimating the Linear Probability Function" *Econometrica*, 36, 1970, pp.775-776.

- McGILLIVRAY, R.G. "Binary Choice of Urban Transport Mode in the San-Francisco Bay Region" *Econometrica*, 40, 1972, pp.827-846.
- NERLOVE M. et PRESS S.J. *Univariate and Multivariate Log-Linear and Logistic Models, Santa-Monica, Calif.* RAND Corporation Report, R.1306, 1973.
- NERLOVE M. et PRESS S.J. *Multivariate Log-Linear Probability Models for the Analysis of Qualitative Data*, Northwestern University, Center for Statistics and Probability, 1976.
- PRAIS S.J., HOUTHAKKER H.S. *The Analysis of Family Budgets*, Cambridge University Press, 1955.
- SCHMIDT, P. et R.P. STRAUSS "Estimation of Models with Jointly Dependent Qualitative Variables : A Simultaneous Logit Approach" *Econometrica*, 43, 1975, pp.745-755.
- THEIL, H. "A Multinomial Extension of the Linear Logit Model", *International Economic Review* 10,1969, pp.251-259.
- THEIL, H. "On the Estimation of Relationships Involving Qualitative Variables", *American Journal of Sociology*, 76, 1970, pp.103-154.
- THEIL, H. *Principles of Econometrics*, Wiley, New-York, 1971.
- TOBIN, J. "Estimation of Relationships for limited Dependent Variables", *Econometrica*, 26, 1958, pp.24-36.
- WALKER, H. et DUNCAN B. "Estimation of the Probability of an Event as a Function of several Independents Variables", *Biometrika*, 54, 1967, pp.167-179
- ZELLNER, A. et LEE T.H. "Joint Estimation of Relationships involving Discrete Random Variables". *Econometrica*, 33, 1965, pp.382-394.

| | |
|---|----|
| . CHAPITRE V : LE MODELE LOGISTIQUE D'EQUATIONS SIMULTANÉES A VARIABLES ENDOGENES QUALITATIVES | 52 |
| V.1. Présentation du modèle | 53 |
| V.2. Estimation du modèle | 55 |
| V.3. Généralisation au cas polytomique | 58 |
| . CONCLUSION | 61 |

o o o o o