L'Actualité économique

La probabilité d’achat de maisons et le modèle logit

Joseph H. Chung

Volume 50, Number 1, janvier–mars 1974

URI: https://id.erudit.org/iderudit/803035ar
DOI: https://doi.org/10.7202/803035ar

See table of contents

Publisher(s)

HEC Montréal

ISSN

0001-771X (print)
1710-3991 (digital)

Explore this journal

Cite this note

Chung, J. H. (1974). La probabilité d’achat de maisons et le modèle logit. *L'Actualité économique*, *50*(1), 88–95. https://doi.org/10.7202/803035ar

*La probabilité d'achat de maisons et le modèle logit**

Le but de cette étude est, d'une part, d'analyser les facteurs motivationnels de l'achat de maisons et, d'autre part, de formuler un modèle probabiliste véritable permettant de prévoir en fonction de certaines caractéristiques socio-économiques des ménages, la proportion des ménages propriétaires.

L'intérêt d'une telle étude est de mieux comprendre le comportement des ménages et de mieux planifier la gestion des entreprises et la politique gouvernementale en matière d'habitation.

— I —

La méthode usuelle de calcul de la probabilité déterminée par plusieurs variables a consisté à mettre en corrélation la variable dépendante binaire avec un ensemble de variables explicatives. L'étude d'Orcutt et de ses collaborateurs [1] portant sur la probabilité d'endettement en est un exemple. La faiblesse majeure de cette méthode est que la valeur estimée de la variable dépendante, qui est supposément la probabilité, varie entre $-\infty$ et $+\infty$ alors que la véritable probabilité doit être comprise entre 0 et 1. Heureusement, une transformation du modèle permet d'éviter cette difficulté.

Deux types de transformation du modèle s'offrent : la transformation probit et la transformation logit. Le modèle probit, qui a été appliqué par certains auteurs [2], est relativement compliqué au point de vue de son fondement mathématique et de l'estimation des para-

---

* Cette étude a été financée par le ministère de l'Education du Québec au cours de l'année 1971-72. Nous tenons à remercier M. Marcel Dagenais, alors directeur du Centre d'Econométrie de l'Institut d'Economie appliquée de l'Ecole des Hautes Etudes commerciales de Montréal, pour son encouragement. Nous remercions d'une façon particulière M. Edouard St-Louis pour sa collaboration technique.

1. Orcutt, Guy H., et al., *Microanalysis of Socio-economic System : A Simulation Study,* Harper & Row, New York, 1961.

2. Finney, D.J., *Probit Analysis : A Statistical Treatment of the Sigmaid Response Curve,* 2e édition, Cambridge University Press, 1962 ; Lisco, T.E., *The Value of Communities Travel Time : A Study in Urban Transportation Study,* Ph. D. Thesis, University of Chicago Press, 1967.

mètres. Par contre, le modèle logit[3] offre un certain avantage sur le plan de la technique d'estimation des paramètres et son fondement mathématique est relativement simple.

Supposons que la probabilité conditionnelle d'achat de maisons (Y) soit fonction de l'âge du chef de ménage, de sa profession et de plusieurs autres variables représentant les caractéristiques socio-économiques des ménages, soit :

$$Y_t = f(X_{1t}, X_{2t}, ..., X_{kt}) \qquad t = 1, 2, ... n \tag{1}$$

où :

$$Y_t \begin{cases} = 1 & \text{si le ménage est propriétaire} \\ = 0 & \text{si le ménage est locataire} \end{cases}$$

L'estimation de l'équation (1) par la technique des moindres carrés ordinaires donne la probabilité conditionnelle, mais il faut remarquer que la valeur estimée de Y varie entre $-\infty$ et $+\infty$, ce qui ne correspond point à la définition de la probabilité. Le modèle logit consiste à transformer la variable Y en probabilité véritable (P) et, ensuite, à définir une nouvelle variable dépendante (logit) qui sera alors régressée sur le même ensemble des variables explicatives, soit :

$$\log(P / 1 - P) = f(X_{1t}, X_{2t}, ..., X_{kt}) \tag{2}$$

La solution de (2) en P donne :

$$P = 1 / [1 + e^{-f(X_{1t}, X_{2t}, ..., X_{kt})}] \tag{3}$$

Trois techniques d'estimation des paramètres s'offrent. Tout d'abord, les paramètres peuvent être estimés à l'aide de la méthode du maximum de vraisemblance[4]. Il s'avère que la convergence de ce type d'estimation est plutôt lente. La deuxième méthode est offerte par Theil[5] et la troisième, par le modèle du CERAU[6] (modèle Bireg).

Le modèle de Theil se résume comme suit :

$$\log(P_{ik} / 1 - P_{ik}) = \beta_0 + \beta_j + \gamma_k \qquad \begin{aligned} i &= 1, 2, ..., n \\ j, k &= 1, 2 \end{aligned} \tag{4}$$

Theil applique ce modèle en vue d'étudier la probabilité de l'entreprise de réviser son plan de production. Les paramètres β_0, β_j et γ_k repré-

3. Theil, Henri, *Principles of Econometrics,* John Wiley and Sons, New York, 1971, chapitre 12.

4. Craig, J.G. et Uhler, R.S., « The Demand for Automobiles », *Discussion Paper No. 27,* Department of Economics, University of British Columbia, 1969 ; Stopher, P.R., « A Probability Model of Travel Code Choice for the Work Journey », Highway Research Record 283, 1969.

5. Theil, Henri, *op. cit.*

6. Rochefort, P.L., Notes de séminaire, Ecole des Hautes Etudes commerciales, décembre 1969.

sentent différentes situations relativement à des « surprises » et aux inventaires. C'est ainsi que, par exemple, (β_1, γ_1) représentent la situation où la surprise est positive et l'inventaire est peu élevé. De même, les paramètres (β_2, γ_2) représentent la situation où la surprise est négative et l'inventaire est considérable.

Ensuite, Theil calcule la proportion observée des entreprises correspondant aux différentes situations et ceci mène au modèle classique de régression, soit :

$$Y = X\beta + U \tag{5}$$

Theil suggère d'estimer les paramètres par la technique des moindres carrés pondérés, le facteur de pondération étant l'écart-type de U qui se définit ainsi :

$$(\sigma_U)^{1/2} = [N_{jk} f_{jk}(1 - f_{jk})]^{1/2}$$

où :

N_{jk} = la fréquence absolue correspondant à la situation jk
f_{jk} = la fréquence relative correspondant à la situation jk

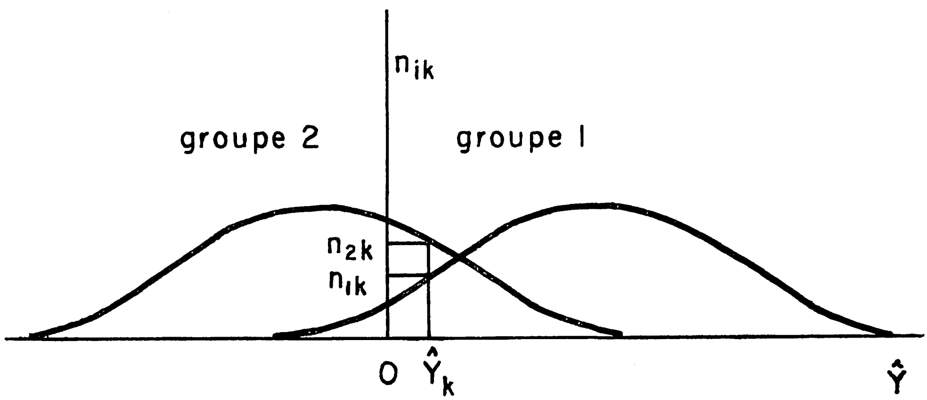
Il est évident que la probabilité du modèle de Theil dépend largement du système de découpage, c'est-à-dire de la façon de définir différentes situations alternatives. Quoi qu'il en soit, la méthode de Theil offre un avantage considérable dans le sens qu'elle implique une méthode de calcul fort simple et qu'elle permet d'effectuer le test de signification statistique des paramètres.

La troisième méthode est offerte par le modèle du CERAU. La méthode consiste, tout d'abord, à effectuer la première régression à l'aide de la technique des moindres carrés ordinaires, soit :

$$\hat{Y} = \hat{f}\ (X_{1t}, X_{2t}, \ldots, X_{kt}) \tag{6}$$

GRAPHIQUE 1

FONCTION DISCRIMINANTE DE $\hat{Y}$



On sait que la fonction de régression qui fait appel à une variable dépendante binaire est proportionnelle à la fonction discriminante [7]. Par conséquent, on peut distribuer $\hat{Y}$ en deux groupes et compter la fréquence absolue correspondant à chaque valeur de $\hat{Y}$.

Au graphique 1, on envisage la distribution normale des deux groupes, un pour le groupe de succès (celui des propriétaires dans notre étude), et l'autre, pour le groupe de non-succès. On y observe n_{1k}, soit la fréquence absolue des propriétaires correspondant à la valeur $\hat{Y}_k$ estimée et n_{2k}, soit la fréquence absolue du deuxième groupe correspondant également à $\hat{Y}_k$.

On définit la probabilité, P, soit :

$$P = \hat{n}_{1k} \,/\, (\hat{n}_{1k} + \hat{n}_{2k}) \qquad (7)$$

On définit ensuite la fonction logistique, à savoir :

$$P \,/\, 1 - P = e^{g(\hat{Y})}$$

ou :

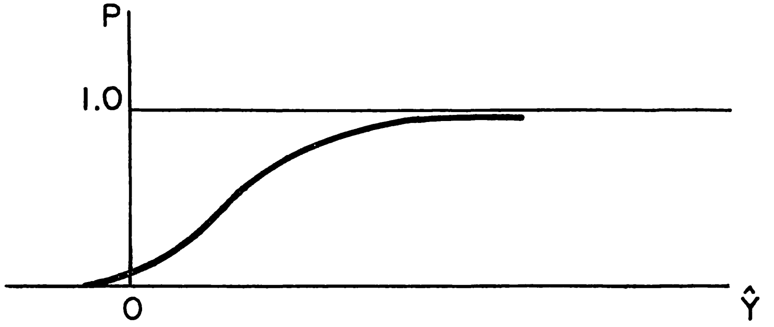
$$P = e^{g(\hat{Y})} \,/\, 1 + e^{g(\hat{Y})} \qquad (8)$$

ce qui donne la courbe logistique indiquée au graphique 2.
Si les deux distributions de $\hat{Y}$ étaient parfaitement normales et avaient la même variance, on aurait une courbe parfaitement logistique.

De (8), on déduit :

$$\begin{aligned} \log\,(P/1 - P) &= g(\hat{Y}) \\ &= g\ [\hat{f}\ (X_{1t}, X_{2t}, ..., X_{kt})] \qquad (9) \end{aligned}$$

GRAPHIQUE 2
FONCTION LOGISTIQUE



7. Rochefort, *op. cit.*

et on peut estimer la probabilité en effectuant la deuxième régression à l'aide de la technique de régression ordinaire.

La probabilité du modèle peut être mise à l'épreuve en comparant la probabilité estimée avec la fréquence relative observée du succès. La probabilité estimée est $\sum_{k=1}^{n} \hat{P}_k$, alors que la fréquence observée est, dans notre étude, le nombre réel de propriétaires.

— II —

Nous avons appliqué le modèle Bireg aux données que nous avons recueillies en 1968 à Sherbrooke grâce à un échantillonnage. Après avoir éliminé les réponses non satisfaisantes, nous avions 76 réponses. Les ménages qui ont répondu à notre échantillonnage étaient tous des occupants de nouveaux logements. Le questionnaire comportait quelques questions sur les mobiles de l'achat de logements. Entre autre, nous demandions les raisons pour lesquelles les ménages choisissaient de devenir propriétaires au lieu de devenir locataires. Il est intéressant de noter que la majorité des répondants, soit 72 p.c. des ménages, évoquèrent le nombre et l'éducation des enfants et l'avantage d'être chez soi comme la raison principale de l'achat de maison. Le facteur de revenu fut évoqué comme la raison principale par environ 17 p.c. des ménages. D'autre part, la mobilité géographique fut souvent évoquée comme la raison du choix d'un logement en location comme mode d'occupation.

De toute façon, ce qui se dégage de notre enquête est que la préférence de logement-propriétaire est autant, sinon plus, motivée par les facteurs socio-psychologiques que par le facteur économique. A la lumière des résultats de notre échantillonnage, la première régression est spécifiée de la manière suivante :

$$Y_t = \sum_{i=1}^{k} \beta_i X_{it} + e_t \qquad \begin{array}{l} i = 1, 2, 3, ..., k \\ t = 1, 2, 3, ..., n \end{array} \qquad (10)$$

où :

Y_t $\begin{cases} = 1 & \text{si le ménage est propriétaire} \\ = 0 & \text{si le ménage est locataire} \end{cases}$

X_{1t} = la mobilité géographique mesurée par la fréquence de déplacement depuis cinq ans

X_{2t} = l'âge du chef de ménage

X_{3t} = le nombre d'années écoulées depuis le mariage

X_{4t} = le nombre d'enfants

$$X_{5t} \begin{cases} = 1 & \text{profession libérale} \\ = 0 & \text{autres} \end{cases}$$

$$X_{6t} \begin{cases} = 1 & \text{métiers spécialisés} \\ = 0 & \text{autres} \end{cases}$$

X_{7t} = l'épargne bancaire du chef de ménage

X_{8t} = les dépenses antérieures d'habitation y compris les dépenses d'ameublement et toutes autres dépenses attribuables à l'habitation.

Les résultats statistiques sont résumés ci-dessous :

$$\begin{aligned} \hat{Y}_t &= \underset{}{-0.6659} - \underset{(2.29)}{0.297\,(X_{1t})} - \underset{(1.20)}{0.0086\,(X_{2t})} \\ &\quad + \underset{(2.72)}{0.019\,(X_{3t})} + \underset{(3.41)}{0.086\,(X_{4t})} + \underset{(2.10)}{0.220\,(X_{5t})} \\ &\quad + \underset{(3.20)}{0.320\,(X_{6t})} - \underset{(2.52)}{0.0011\,(X_{7t})} + \underset{(6.01)}{0.0078\,(X_{8t})} \\ \bar{R}^2 &= 0.605 \qquad N = 76 \end{aligned} \tag{11}$$

D'après ces résultats, les huit variables en tant que groupe expliquent 60 p.c. de la variation totale de la probabilité conditionnelle. Chacune de ces variables, à l'exception de l'âge du chef de ménage, est fortement significative au seuil critique de 5 p.c. de probabilité. Les résultats suggèrent que la probabilité conditionnelle d'achat de maison varie directement avec le nombre d'années écoulées depuis le mariage, le nombre d'enfants, la profession libérale et le métier spécialisé. Ces résultats confirment ceux des études antérieures [8].

L'importance du nombre d'enfants s'explique par le fait connu que les logements en location qui sont suffisamment grands sont plutôt rares au Canada. Que le nombre d'années écoulées depuis le mariage se révèle significatif reflète le fait que l'accumulation des épargnes nécessaires au paiement comptant ne se réalise pas du jour au lendemain. Puisque la contrainte zéro est mise sur le métier non spécialisé, les résultats concernant la profession libérale et le métier spécialisé impliquant que les chefs de ménage ayant un métier spécialisé ou professionnel préfèrent, *ceteris paribus*, être propriétaires de leur logement davantage

8. Rapkin, C. et al., *Housing Market Analysis*, Washington, D.C., 1953 ; Cullingworth, J.B., « Housing Preference », *Town and Country Planning*, vol. 33, 1965 ; Weiss, S.F. et al., « Consumer Preference in Residential Location : A Preliminary Investigation of House Purchase Decision », *Research Review*, vol. 13, 1966.

que ceux qui ont un métier non spécialisé. Ceci peut signifier que l'achat de maison reflète le mode de vie propre à chaque groupe de métiers.

D'autre part, la probabilité d'achat de maison est corréliée à la mobilité géographique et au compte en banque. Il est évident que le déplacement fréquent d'un endroit à l'autre ne se prête guère à l'achat de maison à cause du coût additionnel que chaque achat implique. Quant à la variable compte en banque, le résultat semble signifier qu'on utilise ses propres épargnes en plus de prêts hypothécaires très généreux. Les dépenses d'habitation apparaissent positivement reliées à la probabilité d'achat de maison. Ceci veut dire qu'au fur et à mesure que le coût d'habitation augmente, le ménage trouverait un point critique où l'avantage net d'être chez soi devient positif.

Il faut enfin noter que la variable de revenu ne se révèle pas significative. Il y a peut-être deux explications. D'une part, le revenu courant que nous avons utilisé ne reflète pas vraiment le comportement des ménages en rapport avec l'achat de maisons ; il est bien connu que c'est le revenu permanent et non le revenu courant qui joue le rôle décisif dans la demande de logements [9]. En second lieu, il se peut que la profession reflète une partie au moins du rôle du revenu dans l'achat d'une maison.

Dans l'ensemble, les résultats statistiques sont acceptables et confirment les résultats des études antérieures et les opinions de ceux qui connaissent bien le marché des logements.

La transformation des résultats de la première régression en fonction discriminante est basée sur les hypothèses suivantes :

$$\hat{Y}_1 = N[(\hat{Y}_1 = 0.6983),\ (\sigma\hat{Y}_1 = 0.357)] \qquad (12)$$

$$\hat{Y}_2 = N[(\hat{Y}_2 = 0.1349),\ (\sigma\hat{Y}_2 = 0.209)] \qquad (13)$$

La deuxième régression mène aux résultats suivants :

$$\log\ (P_k/1 - P_k) = 8.74 - 3.67\,\hat{Y}_k$$

ou

$$P_k = (e^{8.74-3.67\,\hat{Y}_k})\ /\ (1 + e^{8.74-3.67\,\hat{Y}_k}) \qquad (14)$$

De plus, nous avons obtenu :

$$\sum_{i=1}^{76} P_{ki} = 27.0735$$

alors que le nombre réel des propriétaires est de 27. C'est ainsi que la transformation logistique de la probabilité conditionnelle assure une fiabilité prévisionnelle très acceptable.

9. Chung, Joseph H., « L'analyse de la demande de logements propriétaires : l'expérience canadienne », *L'Actualité Economique*, vol. 43, n° 1, 1967, pp. 66-86.

— III —

Si le modèle Bireg se révèle utile et démontre une certaine fiabilité prévisionnelle, il n'est pas dépourvu de certaines faiblesses. En premier lieu, il n'existe pas de moyen permettant d'effectuer le test statistique pour les paramètres de la deuxième régression, car la variable dépendante, soit la variable logit, est une variable estimée en vertu de la transformation logistique. Deuxièmement, puisqu'il est probable que la proportion $N_{1k} / (N_{1k} + N_{2k})$ tende à se concentrer près de 1 ou 0, la valeur logarithmique risque d'être zéro ou l'infini, ce qui est naturellement absurde. Dans ce cas, comme nous l'avons fait dans notre étude, il convient de remplacer la proportion observée par celle qui est estimée dans l'hypothèse de distribution normale. Cependant, dans la mesure où l'hypothèse de distribution normale n'est pas exacte, l'estimation du modèle implique des jugements arbitraires.

Quoi qu'il en soit, le modèle Bireg conduit à des résultats supérieurs aux autres modèles logit surtout dans les cas où les variances de deux groupes ne sont pas égales, ce qui est fréquent. Dans ce sens, le modèle Bireg est moins restrictif que d'autres modèles.

En terminant on peut aisément prévoir l'application du modèle dans des études du choix des ménages entre différents modes de transport, de la prévision du choix des entreprises entre différentes décisions et, plus généralement, dans les études de la prise de décision.

Joseph H. CHUNG,
Université du Québec à Montréal.