COURS ECONOMETRIE DES VARIABLES QUALITATIVES

INTRODUCTION

1. Autres noms :

- **Modèles de choix discrets**
- **Modèles à variables endogènes qualitatives (et/ou limitées)**

2. Historique :

L'étude des modèles décrivant les modalités prises par une ou plusieurs variables qualitatives date des années 1940-1950. Les travaux les plus marquants de cette époque sont ceux de Berkson (1944, 1951) consacrés notamment aux modèles dichotomiques simples (modèles logit et probit).

Les premières applications ont été essentiellement menées dans le domaine de la biologie, de la sociologie et de la psychologie. Ce n'est que récemment, que ces modèles ont été utilisés pour décrire des données économiques avec notamment les travaux de Daniel L. MacFadden (1974) et de James J. Heckman (1976). Or, l'application des techniques économétriques propres aux variables qualitatives à des problématiques économiques a d'une part largement contribué à améliorer

l'interprétation des modèles simples (comme par exemple le modèle logit avec les travaux de MacFadden), et d'autre part à identifier des problèmes économiques dont la structure, si elle n'est pas qualitative au sens propre du terme, en est mathématiquement très proche (c'est par exemple le cas de la consommation de bien durable avec le modèle de Tobin de 1958).

Ces développements ont ainsi conduit à introduire un modèle intermédiaire entre les modèles qualitatifs et le modèle linéaire habituel : le modèle tobit

Un des développements majeurs de l'économétrie dans les années 60 et 70, fut lié à l'utilisation croissante des données microéconomiques relatives à des caractéristiques économiques d'agents individuels (firmes, consommateurs, centres de profits…). A cette époque, les bases de données microéconomiques ont en effet pu être constituées, puis exploitées principalement du fait de l'extension des capacités informatiques et de la réduction de leur coût. Bien souvent, les données statistiques disponibles dans ces bases sont relatives à des caractères qualitatifs comme par exemple la catégorie socio-professionnelle, le type d'études suivies, le fait de travailler ou au contraire d'être au chômage, d'acheter ou de ne pas acheter un certain produit etc… Mais les méthodes d'inférence traditionnelles ne permettent pas de modéliser et d'étudier des caractères qualitatifs : des méthodes spécifiques doivent être utilisées tenant compte par exemple de l'absence de continuité des variables

traitées ou de l'absence d'ordre naturel entre les modalités que peut prendre le caractère qualitatif. Les méthodes spécifiques sont les suivantes :

- **modèles à choix binaire**
- **données censurées**
- **logit conditionnel**
- **modèles de comptage**
- **modèles de probabilité linéaire**
- **variable aléatoire logistique**
- **logit**
- **fonction de vraisemblance**
- **effet marginal**
- **modèles à choix multiples**
- **logit multinomial**
- **odds ratio**
- **modèles à choix ordonnés**
- **probit ordonnés**
- **variables ordinales**
- **variable aléatoire de Poisson**
- **modèle de régression de Poisson**
- **probit**
- **biais de sélection**
- **model Tobit**
- **probabilitées**
- **données tronquées…**

Concrètement la variable endogène peut prendre les valeurs suivantes :

1- **Données de comptage : y = 0,1,2…Nombre de voitures**
2- **Choix qualitatifs : y = 0,1. Echec ou réussite. Simple codage 0/1.**
3- **Nombres classés : 0,1,2,3… 0 représente « pas satisfait », 1 « satisfait »…Les valeurs (0 1 2 3) n'ont pas de signification, elle permettent seulement d'ordonner les réponses.**
4- **Codage des CSP : pas de signification et pas d'ordre.**

5

References

- Abramowitz, M. and Stegun, A. (1970), *Handbook of Mathematical Functions*, New York: Dover Press.
- Alban T. (2000), "Econométrie des Variables Qualitatives", Dunod.
- Amemiya T. (1976), "The ML, the Minimum Chi-*Square and the Non Linear Weighted Least Squares Estimator in the General Qualtitative Response Model", Journal of the American Statistical Association, 71, 347-351
- Amemiya T. (1985), "Advanced Econometrics", Cambridge, Harvard University Press.
- Amemiya, T. (1981), "Qualitative Response Models: A Survey," *Journal of Economic Literature*, 19, 483–536.
- Ben-Akiva, M. and Lerman, S. R. (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press Series in Transportation Studies, Cambridge, MA: MIT Press.
- Berkson J. (1944), "Application of the Logistique Function to Bio-Assay", JASA, 39, 357-365.
- Berkson J. (1951), "Why I prefer Logit to Probit", Biometrics, 7, 327-339.
- Bhat, C. R. (1995), "A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice," *Transportation Research*, 29, 471–483.
- Brownstone, D. and Small, K. A. (1989), "Efficient Estimation of Nested Logit Models," *Journal of Business and Economic Statistics*, 7, 67–74.
- Brownstone, D. and Train, K. (1999), "Forecasting New Product Penetration with Flexible Substitution Patterns," *Journal of Econometrics*, 89, 109–129.
- Colletaz G. (2001), "Modèles à Variables Expliquées Qualitatives", Miméo Université Orléans
- Daganzo, C. (1979), *Multinomial Probit: The Theory and Its Application to Demand Forecasting*, New York: Academic Press.
- Daganzo, C. and Kusnic, M. (1993), "Two Properties of the Nested Logit Model," *Transportation Science*, 27, 395–400.
- Davidson R. et MacKinnon J.G. (1984), "Convenient Tests for Logit and Probit Models", Journal of Econometrics, 25, 241-262.
- Estrella, A. (1998), "A New Measure of Fit for Equations with Dichotomous Dependent Variables," *Journal of Business and Economic Statistics*, 16, 198–205.
- Gallant, A. R. (1987), *Nonlinear Statistical Models*, New York: John Wiley & Sons.
- Geweke, J. (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–1339.
- Geweke, J., Keane, M. P., and Runkle, D. (1994), "Alternative Computational Approaches to Inference in the Multinomial Probit Model," *Review of Economics and Statistics*, 76, 609–632.
- Godfrey, L. G. (1988), *Misspecification Tests in Econometrics*, Cambridge: Cambridge University Press.
- Gourieroux C. (1989), "Econométrie des Variables Qualitatives", Economica.
- Gourieroux C. et Montfort A. (1981), "Asymptotic Properties of the Maximuùm Likelihood Estimator in Dichotomous Logit Models", Journal of Econometrics, 17, 83-97.
- Greene, W. H. (1997), *Econometric Analysis*, 3rd Edition, Upper Saddle River, NJ: Prentice-Hall.
- Gregory, A. W. and Veall, M. R. (1985), "On Formulating Wald Tests for Nonlinear Restrictions," *Econometrica*, 53, 1465–1468.

- Hajivassiliou, V. A. (1993), "Simulation Estimation Methods for Limited Dependent Variable Models," in G. S. Maddala, C. R. Rao, and H. D. Vinod, eds., *Econometrics*, volume 11 of *Handbook of Statistics*, New York: Elsevier Science.
- Hajivassiliou, V. A., McFadden, D. L., and Ruud, P. A. (1996), "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results," *Journal of Econometrics*, 72, 85–134.
- Hausman, J. A. and McFadden, D. (1984), "Specification Tests for the Multinomial Logit Model," *Econometrica*, 52, 1219–1240.
- Hensher, D. A. (1986), "Sequential and Full Information Maximum Likelihood Estimation of a Nested Logit Model," *Review of Economics and Statistics*, 68, 657–667.
- Hensher, D. A. (1993), "Using Stated Response Choice Data to Enrich Revealed Preference Discrete Choice Models," *Marketing Letters*, 4, 139–151.
- Judge G.G., Miller D.J. et Mittelhammer R.C. (2000), "Econometric Foundations", Cambridge University Press.
- Keane, M. P. (1994), "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, 62, 95–116.
- Keane, M. P. (1997), "Current Issues in Discrete Choice Modeling," *Marketing Letters*, 8, 307–322.
- Klein R.W. et Spady R.H. (1993), "An Efficient Semi Parametric Estimator for Binary Response Models", Econometrica, 61, 387-421
- LaMotte, L. R. (1994), "A Note on the Role of Independence in *t* Statistics Constructed from Linear Statistics in Regression Models," *American Statistician*, 48, 238–240.
- Long, J. S. (1997). Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage Publications.
- Luce, R. D. (1959), *Individual Choice Behavior: A Theoretical Analysis*, New York: John Wiley & Sons.
- McDonald, J. F. and Moffitt, R. A. 1980. The Uses of Tobit Analysis. The Review of Economics and Statistics Vol 62(2): 318-321.
- Maddala, G. S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, New York: Cambridge University Press.
- McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., *Frontiers in Econometrics*, New York: Academic Press.
- McFadden, D. (1978), "Modelling the Choice of Residential Location," in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds., *Spatial Interaction Theory and Planning Models*, Amsterdam: North-Holland.
- McFadden, D. (1981), "Econometric Models of Probabilistic Choice," in C. F. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MA: MIT Press.
- McFadden, D., Talvitie, A. P., and Associates (1977), *The Urban Travel Demand Forecasting Project: Phase I Final Report Series*, volume 5, Berkeley: Institute of Transportation Studies, University of California, Berkeley.
- McFadden, D. L. and Ruud, P. A. (1994), "Estimation by Simulation," *Review of Economics and Statistics*, 76, 591–608.
- Morimune K. (1979), "Comparisons of Normal and Logistic Models in the Bivariate Dichitomous Analysis", Econometrica, 47, 957-975.
- Phillips, C. B. and Park, J. Y. (1988), "On Formulating Wald Tests of Nonlinear Restrictions," *Econometrica*, 56, 1065–1083.

- Powers, D. A. and Xie, Y. (2000), *Statistical Methods for Categorical Data Analysis*, San Diego: Academic Press.
- Radner R. et Miller L. (1970), "Demand and Supply in U.S. Higher Education : A Progress Report", American Economic Review, 60.
- Schmidt, P. and Strauss, R. (1975), "The Prediction of Occupation Using Multiple Logit Models," *International Economic Review*, 16, 471–486.
- Small, K. (1982), "The Scheduling of Consumer Activities: Work Trips," *American Economic Review*, 72, 467–479.
- Spector, L. and Mazzeo, M. (1980), "Probit Analysis and Economic Education," *Journal of Economic Education*, 11, 37–44.
- Swait, J. and Bernardino, A. (2000), "Distinguishing Taste Variation from Error Structure in Discrete Choice Data," *Transportation Research Part B*, 34, 1–15.
- Theil, H. (1969), "A Multinomial Extension of the Linear Logit Model," *International Economic Review*, 10, 251–259.
- Tobin J. (1958), "Estimation of Relationships for Limited Dependent Variables", Econometrica, 26, 24-36
- Train, K. E., Ben-Akiva, M., and Atherton, T. (1989), "Consumption Patterns and Self-Selecting Tariffs," *Review of Economics and Statistics*, 71, 62–73

References (par ordre chronologique)

- Berkson J. (**1944**), ”Application of the Logistique Function to Bio-Assay”, JASA, 39, 357-365.
- Berkson J. (**1951**), ”Why I prefer Logit to Probit”, Biometrics, 7, 327-339.
- Abramowitz, M. and Stegun, A. (1970), *Handbook of Mathematical Functions*, New York: Dover Press.
- Tobin J. (**1958**), ”Estimation of Relationships for Limited Dependent Variables”, Econometrica, 26, 24-36
- Luce, R. D. (**1959**), *Individual Choice Behavior: A Theoretical Analysis*, New York: John Wiley & Sons.
- Theil, H. (**1969**), “A Multinomial Extension of the Linear Logit Model,” *International Economic Review*, 10, 251–259.
- Radner R. et Miller L. (**1970**), ”Demand and Supply in U.S. Higher Education : A Progress Report”, American Economic Review, 60.
- McFadden, D. (**1974**), “Conditional Logit Analysis of Qualitative Choice Behavior,” in P. Zarembka, ed., *Frontiers in Econometrics*, New York: Academic Press.
- Schmidt, P. and Strauss, R. (**1975**), “The Prediction of Occupation Using Multiple Logit Models,” *International Economic Review*, 16, 471–486.
- Amemiya T. (**1976**), ”The ML, the Minimum Chi-*Square and the Non Linear Weighted Least Squares Estimator in the General Qualtitative Response Model”, Journal of the American Statistical Association, 71, 347-351
- McFadden, D., Talvitie, A. P., and Associates (**1977**), *The Urban Travel Demand Forecasting Project: Phase I Final Report Series*, volume 5, Berkeley: Institute of Transportation Studies, University of California, Berkeley.
- McFadden, D. (**1978**), “Modelling the Choice of Residential Location,” in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds., *Spatial Interaction Theory and Planning Models*, Amsterdam: North-Holland.
- Daganzo, C. (**1979**), *Multinomial Probit: The Theory and Its Application to Demand Forecasting*, New York: Academic Press.
- Morimune K. (**1979**), ”Comparisons of Normal and Logistic Models in the Bivariate Dichitomous Analysis”, Econometrica, 47, 957-975.
- Spector, L. and Mazzeo, M. (**1980**), “Probit Analysis and Economic Education,” *Journal of Economic Education*, 11, 37–44.
- McDonald, J. F. and Moffitt, R. A. (**1980**). The Uses of Tobit Analysis. The Review of Economics and Statistics Vol 62(2): 318-321.
- Amemiya, T. (**1981**), “Qualitative Response Models: A Survey,” *Journal of Economic Literature*, 19, 483–536.
- Gourieroux C. et Montfort A. (**1981**), ”Asymptotic Properties of the Maximuùm Likelihood Estimator in Dichotomous Logit Models”, Journal of Econometrics, 17, 83-97.
- McFadden, D. (**1981**), “Econometric Models of Probabilistic Choice,” in C. F. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MA: MIT Press.
- Small, K. (**1982**), “The Scheduling of Consumer Activities: Work Trips,” *American Economic Review*, 72, 467–479.
- Maddala, G. S. (**1983**), *Limited-Dependent and Qualitative Variables in Econometrics*, New York: Cambridge University Press.

- Hausman, J. A. and McFadden, D. (**1984**), "Specification Tests for the Multinomial Logit Model," *Econometrica*, 52, 1219–1240.
- Davidson R. et MacKinnon J.G. (**1984**), "Convenient Tests for Logit and Probit Models", Journal of Econometrics, 25, 241-262.
- Gregory, A. W. and Veall, M. R. (**1985**), "On Formulating Wald Tests for Nonlinear Restrictions," *Econometrica*, 53, 1465–1468.
- Amemiya T. (**1985**), "Advanced Econometrics", Cambridge, Harvard University Press.
- Ben-Akiva, M. and Lerman, S. R. (**1985**), *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press Series in Transportation Studies, Cambridge, MA: MIT Press.
- Hensher, D. A. (**1986**), "Sequential and Full Information Maximum Likelihood Estimation of a Nested Logit Model," *Review of Economics and Statistics*, 68, 657–667.
- Gallant, A. R. (**1987**), *Nonlinear Statistical Models*, New York: John Wiley & Sons.
- Godfrey, L. G. (1988), *Misspecification Tests in Econometrics*, Cambridge: Cambridge University Press.
- Phillips, C. B. and Park, J. Y. (**1988**), "On Formulating Wald Tests of Nonlinear Restrictions," *Econometrica*, 56, 1065–1083.
- Gourieroux C. (**1989**), "Econométrie des Variables Qualitatives", Economica.
- Train, K. E., Ben-Akiva, M., and Atherton, T. (**1989**), "Consumption Patterns and Self-Selecting Tariffs," *Review of Economics and Statistics*, 71, 62–73
- Geweke, J. (**1989**), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–1339.
- Brownstone, D. and Small, K. A. (**1989**), "Efficient Estimation of Nested Logit Models," *Journal of Business and Economic Statistics*, 7, 67–74.
- Daganzo, C. and Kusnic, M. (**1993**), "Two Properties of the Nested Logit Model," *Transportation Science*, 27, 395–400.
- Klein R.W. et Spady R.H. (**1993**), "An Efficient Semi Parametric Estimator for Binary Response Models", Econometrica, 61, 387-421
- Hensher, D. A. (**1993**), "Using Stated Response Choice Data to Enrich Revealed Preference Discrete Choice Models," *Marketing Letters*, 4, 139–151.
- Hajivassiliou, V. A. (**1993**), "Simulation Estimation Methods for Limited Dependent Variable Models," in G. S. Maddala, C. R. Rao, and H. D. Vinod, eds., *Econometrics*, volume 11 of *Handbook of Statistics*, New York: Elsevier Science.
- McFadden, D. L. and Ruud, P. A. (**1994**), "Estimation by Simulation," *Review of Economics and Statistics*, 76, 591–608.
- LaMotte, L. R. (**1994**), "A Note on the Role of Independence in *t* Statistics Constructed from Linear Statistics in Regression Models," *American Statistician*, 48, 238–240.
- Keane, M. P. (**1994**), "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, 62, 95–116.
- Geweke, J., Keane, M. P., and Runkle, D. (**1994**), "Alternative Computational Approaches to Inference in the Multinomial Probit Model," *Review of Economics and Statistics*, 76, 609–632.
- Bhat, C. R. (**1995**), "A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice," *Transportation Research*, 29, 471–483.
- Hajivassiliou, V. A., McFadden, D. L., and Ruud, P. A. (**1996**), "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results," *Journal of Econometrics*, 72, 85–134.

- Greene, W. H. (**1997**), *Econometric Analysis*, 3rd Edition, Upper Saddle River, NJ: Prentice-Hall.
- Keane, M. P. (**1997**), "Current Issues in Discrete Choice Modeling," *Marketing Letters*, 8, 307–322.
- Long, J. S. (**1997**). Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage Publications.
- Estrella, A. (**1998**), "A New Measure of Fit for Equations with Dichotomous Dependent Variables," *Journal of Business and Economic Statistics*, 16, 198–205.
- Brownstone, D. and Train, K. (**1999**), "Forecasting New Product Penetration with Flexible Substitution Patterns," *Journal of Econometrics*, 89, 109–129.
- Alban T. (**2000**), "Econométrie des Variables Qualitatives", Dunod.
- Judge G.G., Miller D.J. et Mittelhammer R.C. (**2000**), "Econometric Foundations", Cambridge University Press.
- Swait, J. and Bernardino, A. (**2000**), "Distinguishing Taste Variation from Error Structure in Discrete Choice Data," *Transportation Research Part B*, 34, 1–15.
- Powers, D. A. and Xie, Y. (**2000**), *Statistical Methods for Categorical Data Analysis*, San Diego: Academic Press.
- Colletaz G. (**2001**), "Modèles à Variables Expliquées Qualitatives", Miméo Université Orléans