

**PSEUDO-MAXIMUM DE VRAISEMBLANCE : COMMENT
ESTIMER SIMPLEMENT DES MODÈLES QUALITATIFS
STATIQUES AVEC VARIABLES ENDOGÈNES SUR
DONNÉES DE PANEL**

Stéfan Lollivier ()*

()Insee*

Résumé :

Sur données qualitatives, l'intérêt des méthodes de pseudo-maximum de vraisemblance est de permettre d'éviter le recours quasi obligé à des restrictions paramétriques mais également des complexités calculatoires. Ces méthodes remplacent la « vraie » vraisemblance par une vraisemblance approchée avec un modèle en général normal, mais qui respecte les moments conditionnels d'ordre un, voire deux, de la variable expliquée (GOURIÉROUX, MONFORT, TROGNON, [1984]). On peut ainsi facilement montrer que l'estimateur des paramètres des variables explicatives du modèle probit simple empilé est convergent même lorsque la matrice de variance du terme d'erreur est non diagonale, et quelle que soit la forme de cette matrice. Le calcul de la matrice de variance de ces estimateurs est plus complexe, mais on peut facilement l'obtenir grâce à des techniques de bootstrap.

C'est en présence de variables explicatives endogènes que l'utilisation des méthodes de pseudo-maximum de vraisemblance est la plus décisive car l'estimation par la méthode du MV d'un système de variables latentes sur données longitudinales est quasiment impossible. Le problème est déjà complexe lorsque la variable de l'équation d'intérêt est continue, et différents auteurs ont proposé des méthodes alternatives à celles du MV. Ainsi, même en coupe transversale, HECKMAN ([1974, 1978]) a montré que le recours à une procédure en deux étapes était à la fois plus simple à mettre en œuvre, mais aussi plus économe en hypothèses sur les lois des termes d'erreur. WOOLDRIDGE [1995] a généralisé cette approche en deux étapes aux données de panel, mais la variable expliquée demeure quantitative. D'autres auteurs (KYRIAZIDOU [1997], DUSTMANN ROCCHINA-BARRACHINA [2000]) ont mis au point des techniques d'estimation plus économes en hypothèses paramétriques, notamment sur les lois des termes d'erreur, mais au prix de difficultés d'estimation encore accrues.

La situation dans laquelle toutes les équations font intervenir des variables qualitatives fait l'objet d'une littérature est encore peu développée. On décrit ici une méthode en plusieurs étapes, comme chez HECKMAN et WOOLDRIDGE. La première étape consiste à estimer de façon convergente l'équation instrumentale. Dans un deuxième temps, on tire pour chaque individu un résidu dans la loi du terme d'erreur conditionnelle à l'observation. La troisième étape réalise l'estimation de l'équation d'intérêt au moyen d'une méthode de pseudo-maximum de vraisemblance simulé (GOURIÉROUX, MONFORT [1993]). Les estimateurs ainsi obtenus sont alors convergents et asymptotiquement normaux, dès lors que le nombre d'individus et le nombre de simulations tend vers l'infini. L'estimation est assez simple lorsque l'équation instrumentale est univariée, plus complexe, mais réalisable en utilisant une méthode d'échantillonnage de Gibbs, lorsque l'équation instrumentale est multivariée.

JEL classification numbers : C33, C35, C15.

Keywords : panel data, endogenous variables, simulation

Introduction

La prise en compte de l'endogénéité d'une variable explicative dans une équation de comportement conduit fréquemment à expliciter puis à estimer un système d'équations simultanées se présentant sous forme structurelle. La plupart du temps, il en résulte que l'estimation de l'équation d'intérêt sans tenir compte de l'endogénéité conduit à des résultats biaisés. La raison est que, dans l'équation en question, l'espérance du terme d'erreur conditionnelle à la variable explicative décrite dans la seconde équation instrumentale n'est pas nulle.

Le traitement de l'endogénéité des variables explicatives est maintenant bien établi sur données de coupe transversale lorsque les variables expliquées et explicatives sont continues. La méthode des doubles moindres carrés peut même s'appliquer lorsque la variable explicative endogène est qualitative. Parmi les approches disponibles, la méthode des régressions augmentées permet de réaliser aisément des tests de spécification mais aussi d'effectuer des estimations permettant de corriger de tels biais (voir par exemple ROBIN [2000]). Cette méthode consiste à ajouter à la régression une variable supplémentaire le plus souvent construite à partir d'une estimation auxiliaire afin de corriger les biais d'endogénéité. On dispose ainsi de procédures pour traiter ces biais, la difficulté dans la pratique consistant seulement à trouver des instruments adéquats.

Cependant, de plus en plus de travaux appliqués ont recours à des estimations économétriques dans lesquelles la variable expliquée, voire les variables expliquées et explicatives, sont qualitatives. L'estimation du modèle à équations simultanées qui en résulte est alors plus délicate, à fortiori lorsque l'instrumentation renvoie à un modèle polytomique non ordonné. On dispose classiquement de la méthode du maximum de vraisemblance. Elle fournit des estimateurs convergents et permet de tester l'endogénéité de la (ou des) variable explicative dans l'équation d'intérêt. Mais cette technique fait appel à des hypothèses paramétriques lourdes et restrictives, notamment la normalité jointe temporelle des termes d'erreur dans le système d'équations. Il en résulte une écriture de la vraisemblance faisant intervenir des corrélations des termes d'erreur entre eux et dans le temps, et dont le calcul est compliqué et nécessite l'intégration de fonctions normales multidimensionnelles. Le recours à des techniques de simulation est en général incontournable. De même, afin d'alléger les calculs, il est souvent indispensable d'effectuer des hypothèses simplificatrices sur la matrice de variance temporelle des termes d'erreur, ce qui rend le modèle plus sensible aux erreurs de spécification.

Ces difficultés peuvent néanmoins être contournées dans un certain nombre de cas. Par exemple, dans la situation la plus simple où la variable expliquée est continue et la variable explicative endogène dichotomique, HECKMAN [1974, 1978] a montré que le recours à une procédure en deux étapes était à la fois plus commode à mettre en place, mais aussi plus économe en hypothèses sur les lois des termes d'erreur. En outre, une telle approche peut être généralisée.

L'extension des modèles à équations simultanées avec variable d'intérêt quantitative aux données longitudinales est récente, qu'il s'agisse de modèles de sélection, ou de résolution de problèmes d'endogénéité, le formalisme sous-jacent étant analogue. Le modèle développé par WOOLDRIDGE [1995] vise à généraliser la méthode en deux étapes d'HECKMAN aux données longitudinales. Moyennant un jeu d'hypothèses portant essentiellement sur les termes d'erreur, WOOLDRIDGE parvient à construire des estimateurs convergents et asymptotiquement normaux des paramètres d'intérêt, et notamment de ceux de l'équation soumise au biais de sélection. La technique d'estimation est en outre relativement simple, voisine de celle présentée par HECKMAN. D'autres auteurs (KYRIAZIDOU [1997], DUSTMANN, ROCCHINA-BARRACHINA [2000]) ont mis au point des techniques d'estimation plus économes en hypothèses paramétriques, notamment sur les lois des termes d'erreur, mais au prix de difficultés d'estimation accrues.

On se propose ici de décrire comment il est possible de recourir à des techniques de pseudo-maximum de vraisemblance simulé pour estimer en plusieurs étapes des modèles qualitatifs avec endogénéité sur données longitudinales. L'intérêt des méthodes de pseudo-maximum de vraisemblance est de remplacer dans les formules de calcul la loi compliquée du terme d'erreur par une loi plus simple de même moment d'ordre 1, tout en conservant les propriétés de convergence asymptotique (GOURIÉROUX, MONFORT [1993]). On procède en plusieurs étapes : estimation séparée de l'équation d'instrumentation, puis simulation des termes d'erreur dans l'équation d'instrumentation

afin de corriger les biais d'endogénéité dans l'équation d'intérêt ; enfin, estimation de l'équation d'intérêt augmentée comme un pseudo-maximum de vraisemblance simulé. Les estimateurs ainsi obtenus sont alors convergents et asymptotiquement normaux, dès lors que le nombre d'individus et le nombre de simulations tend vers l'infini. Ils sont en outre robustes par rapport aux erreurs de spécification relatifs à la corrélation temporelle entre les termes d'erreur, puisqu'ils ne font intervenir d'hypothèses paramétriques que sur la loi marginale en t des termes d'erreurs

La méthode est particulièrement simple à mettre en œuvre lorsque l'instrumentation renvoie à un modèle polytomique ordonné (LOLLIVIER [2008]). L'estimateur probit empilé est convergent, et la simulation du terme d'erreur porte sur une variable normale tronquée. Elle est plus complexe à réaliser dans le cas d'un modèle polytomique non ordonné, qui fait intervenir plusieurs variables latentes et des termes d'erreur corrélés entre eux dans le système d'équations instrumentales. L'estimateur en coupe est convergent, et disponible dans la plupart des logiciels statistiques. Il est en revanche plus complexe de simuler les termes d'erreurs conditionnellement aux observations. Une solution, assez simple à programmer, consiste à recourir à l'échantillonnage de Gibbs. Il s'agit d'un procédé de mise à jour markovien qui, à partir de valeurs initiales correspondant à des tirages dans une loi a priori, procède à des cycles de mises à jour séquentielles au moyen de tirages dans les lois conditionnelles des variables. Lorsque le nombre de cycles de mise à jour tend vers l'infini, on obtient, sous certaines hypothèses, les tirages recherchés dans la loi jointe de la variable (GEMAN ET GEMAN [1984]).

1. Rappel sur les m-estimateurs

1.1. Notations

On considère que les unités statistiques observées sont décrites comme des individus $i = 1, \dots, N$. Un individu donné peut correspondre à plusieurs observations dans l'échantillon, la plupart du temps repérées par un indice $t = 1, \dots, T$.

Les variables aléatoires sont, sauf exception, représentées par des lettres majuscules, les réalisations au moyen de minuscules. Les observations y_i et x_i sont supposées tirées de façon indépendantes et équidistribuées dans une population décrite par les variables aléatoires Y et X . Si θ représente l'ensemble des paramètres que l'on cherche à estimer, on notera θ_0 la « vraie » valeur.

1.2. Le formalisme des m-estimateurs

Les m-estimateurs sont des estimateurs extrémaux, obtenus suite à la maximisation d'une expression du type :

$$q_N(\theta) = \frac{1}{N} \sum_{i=1}^N \psi(y_i, x_i, \theta).$$

Si $q_N(\theta)$ tend uniformément vers $E\psi(Y, X, \theta)$, alors $\hat{\theta}_N$ converge vers la pseudo-vraie valeur θ_∞ qui maximise $E\psi(Y, X, \theta)$. La condition requise $\theta_\infty = \theta_0$ n'est bien sûr pas satisfaite pour une fonction ψ quelconque. Si la convergence vers θ_0 est assurée, et moyennant des conditions de régularité, $\sqrt{N}(\hat{\theta}_N - \theta_0)$ converge en loi vers $N(0, J_0^{-1} I_0 J_0^{-1})$, où :

$$I_0 = V \frac{\partial \psi}{\partial \theta}(Y, X, \theta_0) \approx \frac{1}{N} \sum_{i=1}^N \frac{\partial \psi}{\partial \theta}(y_i, x_i, \hat{\theta}_N) \frac{\partial \psi}{\partial \theta}(y_i, x_i, \hat{\theta}_N),$$

et

$$J_0 = -E \frac{\partial^2 \psi}{\partial \theta \partial \theta'}(Y, X, \theta_0) \approx -\frac{1}{N} \sum_{i=1}^N \frac{\partial^2 \psi}{\partial \theta \partial \theta'}(y_i, x_i, \hat{\theta}_N).$$

1.3. Le pseudo-maximum de vraisemblance

Dans une famille de fonctions ψ donnée, l'estimateur qui fournit la plus grande précision est celui du maximum de vraisemblance. Soit en effet $f(y, x, \theta)$ la densité du modèle. Si on choisit $\psi = \log f$, l'inégalité de Jensen assure la convergence de l'estimateur. On peut en outre montrer que l'estimateur est asymptotiquement efficace, et que $I_0 = J_0$.

En revanche, l'estimateur du maximum de vraisemblance n'est pas toujours facile à calculer, et il n'est pas robuste : une erreur de spécification le rend non convergent.

D'autres méthodes, comme celle des GMM ou celle des moindres carrés non linéaires utilisées lorsque la variable expliquée est observée sont moins efficaces mais plus robustes. On décrit ici la méthode du pseudo-maximum de vraisemblance, bien adaptée aux variables discrètes (GOURIÉROUX, MONFORT, TROGNON [1984]). L'idée est de remplacer la densité du modèle par celle d'un pseudo-modèle mal spécifié, mais vérifiant des propriétés portant sur les moments de la variable observée, en l'occurrence le moment d'ordre 1.

Soit $m(X, \theta_0) = E(Y|X)$ la vraie espérance conditionnelle de la variable observée, calculée donc avec la densité f . On suppose que l'on ne commet pas d'erreur de spécification sur l'espérance, et que θ est identifiable au premier ordre :

$$P[m(X, \theta) = m(X, \theta_0)] = 1 \Rightarrow \theta = \theta_0.$$

On s'intéresse enfin aux fonctions ψ de la forme $\psi(Y, X, \theta) = \log g(Y, m(X, \theta))$, où $g(Y, m)$ est la vraisemblance d'un pseudo-modèle d'espérance m . On peut alors montrer que l'estimateur $\hat{\theta}_N$ est convergent si et seulement si le pseudo-modèle est linéaire exponentiel :

$$\log g(Y, m) = A(m) + B(Y) + C(m)Y.$$

1.4. Le pseudo-maximum de vraisemblance simulé

Dans certaines situations, l'espérance conditionnelle de la variable observée n'est pas directement calculable, par exemple si son expression incorpore une variable aléatoire inobservée u_i dont la loi est complexe. On peut alors faire apparaître le moment conditionnel à u_i :

$$\hat{m}(x_i, \theta, u_i) = E(y_i | x_i, u_i),$$

et en déduire le simulateur :

$$m(x_i, \theta) \approx \frac{1}{H} \sum_{h=1}^H \hat{m}(x_i, \theta, u_{ih}).$$

Les propriétés asymptotiques du pseudo-maximum de vraisemblance simulé sont analogues à celles du pseudo-maximum de vraisemblance (GOURIÉROUX, MONFORT [1993]). En particulier, si H et N tendant vers l'infini avec \sqrt{N}/H tendant vers 0, l'estimateur du pseudo maximum de vraisemblance simulé est convergent, asymptotiquement normal et de même matrice de variance asymptotique que

l'estimateur du pseudo maximum de vraisemblance. Pour H fixé et N tendant vers l'infini, l'estimateur est biaisé, mais le biais est en $1/H$, ce qui rend l'estimateur tout à fait utilisable sur données individuelles dès lors que H est suffisamment grand.

2. Le modèle probit dichotomique en panel

2.1. L'estimation

Avec un modèle probit dichotomique longitudinal statique, le modèle s'écrit :

$$y_{it}^* = x_{it}\beta + u_{it}, \text{ avec } y_{it} = I(y_{it}^* > 0).$$

Pour un individu, on a matriciellement :

$$y_i^* = x_i\beta + u_i, \text{ où } u_i \sim N(0, \Omega),$$

et l'on observe le vecteur y_i .

En outre, $m_t(x_{it}, \beta) = E(y_{it} | x_{it}) = \Phi(x_{it}\beta)$. Soit $m(x_i, \beta)$ le vecteur des $m_t(x_{it}, \beta)$.

Si l'on choisit $A(m) = \sum_{t=1}^T \log(1 - m_t)$, $B(m) = 0$ et $C(m)' = \begin{bmatrix} \log[m_1 / (1 - m_1)] \\ \vdots \\ \log[m_T / (1 - m_T)] \end{bmatrix}$, alors la pseudo-

vraisemblance s'écrit :

$$q_n(\beta) = \sum_{t=1}^T \log(1 - m_t(x_{it}, \beta)) + \sum_{t=1}^T y_{it} \log \frac{m_t(x_{it}, \beta)}{1 - m_t(x_{it}, \beta)} = \sum_{i=1}^N \sum_{t=1}^T [y_{it} \log \Phi(x_{it}\beta) + (1 - y_{it}) \log(1 - \Phi(x_{it}\beta))]$$

Elle correspond exactement à la vraisemblance d'un modèle probit en panel pour laquelle la matrice de variance-covariance du terme d'erreur Ω serait l'identité. En d'autres termes, la pseudo-vraisemblance est égale à la vraisemblance du modèle probit empilé. On peut également montrer que cet estimateur est un estimateur obtenu avec la méthode des moments conditionnels (BREITUNG ET LECHNER [1996], BERTSCHEK ET LECHNER [1998]).

De ce fait, l'estimateur du maximum de vraisemblance du modèle probit empilé est convergent même lorsque la matrice Ω n'est pas diagonale. Pour l'estimer, il n'est pas besoin d'effectuer d'hypothèses particulières sur cette matrice. Cet estimateur du pseudo-maximum de vraisemblance est donc robuste aux erreurs de spécifications que l'on est susceptible de commettre lorsque l'on effectue des hypothèses sur Ω destinées à simplifier l'estimation du modèle par la méthode du maximum de vraisemblance. Le calcul est en outre beaucoup plus simple, car il n'est pas nécessaire de recourir à des méthodes d'estimation complexes (simulation,...) surtout lorsque la dimension temporelle est importante.

2.2. Le calcul de la matrice de variance-covariance de l'estimateur

Cette méthode ne fournit cependant que des estimateurs des paramètres β . Or, pour mener à bien des tests, et notamment de significativité, il est nécessaire de déterminer un estimateur de la matrice de variance-covariance asymptotique des paramètres. Une façon de les obtenir est de recourir à une

technique de bootstrap par paires, les termes d'erreur étant I.I.D. par rapport aux individus (EFRON, TIBSHIRANI [1986]).

La méthode mise en œuvre consiste à effectuer un tirage avec remise de taille N parmi les N individus (ceux-ci étant indexés par i ; à chaque individu correspond la totalité des T dates). L'estimateur probit empilé sur chaque échantillon bootstrap ainsi constitué fournit des paramètres $\hat{\beta}_b, b=1, \dots, B$ pour les deux équations. Lorsque B tend vers l'infini, la moyenne et la matrice de variance-covariance empiriques des $\hat{\beta}_b, b=1, \dots, B$ sont des estimateurs convergents des paramètres et de leur matrice de variance-covariance. Dans la mesure où l'échantillon est de grande taille, on se situe dans des conditions asymptotiques pour l'estimation des paramètres du modèle ($N \rightarrow +\infty$). Or, on sait que les estimateurs du pseudo-maximum de vraisemblance suivent asymptotiquement une loi normale. Il n'est donc pas nécessaire d'utiliser la technique du bootstrap pour approximer la loi asymptotique des paramètres, puisqu'elle est connue¹. Le bootstrap est en revanche précieux pour disposer d'un estimateur de la matrice de variance-covariance asymptotique des paramètres, et conduire les tests.

3. Endogénéité dans un modèle bivarié

3.1. Le modèle de Wooldridge, extension du modèle d'Heckman

Le modèle d'Heckman est un modèle bivarié en coupe qui comporte une équation d'intérêt quantitative et une équation de sélection qualitative :

$$\begin{cases} y_{i1} &= x_{i1}\beta_1 + \varepsilon_{i1}, & \text{si } y_{i2} = I(y_{i2}^* > 0) = 1 \\ y_{i2}^* &= x_{i2}\beta_2 + \varepsilon_{i2} \end{cases}$$

On suppose en outre que $E(\varepsilon_{i1} | \varepsilon_{i2}) = \mu \varepsilon_{i2}$. Le terme d'erreur de la deuxième équation est normal, centré et réduit, et on montre aisément que :

$$E(y_{i1} | y_{i2} = 1) = x_{i1}\beta_1 + \mu \frac{\varphi(x_{i2}\beta_2)}{\Phi(x_{i2}\beta_2)}.$$

L'estimation séparée de l'équation de sélection fournit un estimateur convergent de β_2 . Dès lors que N tend vers l'infini, l'estimation de l'équation :

$$y_{i1} = x_{i1}\beta_1 + \mu \frac{\varphi(x_{i2}\hat{\beta}_2)}{\Phi(x_{i2}\hat{\beta}_2)} + \eta_i$$

par les moindres carrés ordinaires fournit un estimateur convergent et asymptotiquement normal de β_1 , dont la matrice de variance-covariance peut être estimée par exemple par bootstrap.

On notera que cette méthode d'estimation ne requiert pas d'hypothèse particulière sur la loi de ε_{i1} , notamment de normalité, contrairement à celle du maximum de vraisemblance, ce qui lui procure une plus grande robustesse.

Le modèle de Wooldridge est une extension du modèle d'Heckman au cas dynamique :

¹ Il serait en revanche légitime de recourir au bootstrap pour approximer la loi des paramètres à distance finie, car celle-ci est inconnue.

$$\begin{cases} y_{it1}^* &= x_{it1}\beta_1 + \varepsilon_{it1}, & \text{si } y_{it2} = I(y_{it2}^* > 0) = 1 \\ y_{it2}^* &= x_{it2}\beta_2 + \varepsilon_{it2} \end{cases}$$

avec toujours $E(\varepsilon_{it1}|\varepsilon_{it2}) = \rho\varepsilon_{it2}$. Dans sa version la plus simple, le terme d'erreur de l'équation de sélection est centré et réduit à toutes les dates.

Une première étape consiste à estimer le modèle de sélection comme un modèle probit empilé. On a vu précédemment que ceci permettait d'obtenir un estimateur convergent de β_2 quelque soit la matrice de variance de ε_{i2} . L'équation d'intérêt augmentée est alors estimée par les moindres carrés ordinaires, ce qui fournit un estimateur convergent de β_1 relativement robuste par rapport aux hypothèses de spécification sur le terme d'erreur joint $(\varepsilon_{i1}, \varepsilon_{i2})$.

3.2. Une méthode en trois étapes

L'idée est d'appliquer la démarche précédente en plusieurs étapes aux modèles purement qualitatifs. On s'intéresse dans cette section à la situation où la variable expliquée est dichotomique, et la variable explicative polytomique ordonnée (ce qui inclut le cas dichotomique). Ces deux variables renvoient donc à un modèle latent bivarié :

$$\begin{cases} y_{it1}^* &= x_{it1}\beta_1 + \sum_{k=0}^K \alpha_k d_{ik} + \varepsilon_{it1}, \\ y_{it2}^* &= x_{it2}\beta_2 + \varepsilon_{it2}, \end{cases}$$

dans lequel le couple $(\varepsilon_{it1}, \varepsilon_{it2})$ suit une loi normale bi-dimensionnelle. On peut alors écrire :

$$\varepsilon_{it1} = \rho\varepsilon_{it2} + \sqrt{1-\rho^2}w_{it},$$

où w_{it} suit une loi normale centrée et réduite, et est indépendant de ε_{it2} . La première équation s'écrit alors :

$$y_{it1}^* = x_{it1}\beta_1 + \sum_{k=0}^K \alpha_k d_{ik} + \rho\varepsilon_{it2} + \sqrt{1-\rho^2}w_{it}.$$

Intuitivement, si l'on connaissait ε_{it2} , on pourrait estimer cette première équation indépendamment de la seconde, comme un modèle probit simple. Cette remarque peut se formaliser de la façon suivante :

$$E(y_{it1}^* | x_i, d_{ik}, k=0, \dots, K; \varepsilon_{it2}) = \Phi \left(\frac{x_{it1}\beta_1 + \sum_{k=0}^K \alpha_k d_{ik} + \rho\varepsilon_{it2}}{\sqrt{1-\rho^2}} \right),$$

et par conséquent, non conditionnellement à ε_{it2} :

$$E(y_{it1}^* | x_i, d_{ik}, k=0, \dots, K) = E_{\varepsilon_{it2}} \left[\Phi \left(\frac{x_{it1}\beta_1 + \sum_{k=0}^K \alpha_k d_{ik} + \rho\varepsilon_{it2}}{\sqrt{1-\rho^2}} \right) \right].$$

Cette expression fait intervenir l'espérance d'une fonction de la variable aléatoire ε_{it2} conditionnelle à l'observation des $d_{ik}, k=0, \dots, K$. La loi de ε_{it2} conditionnelle aux d_{ik} est une loi compliquée, pour laquelle on n'a pas de représentation analytique simple. En revanche, disposer de tirages $\varepsilon_{it2h}, h=1, \dots, H$, dans cette loi permettrait d'approximer l'espérance ci-dessus par un simulateur :

$$E(y_{it1} | x_i, d_{ik}, k=0, \dots, K) \approx \frac{1}{H} \sum_{h=1}^H \left[\Phi \left(\frac{x_{it1} \beta_1 + \sum_{k=0}^K \alpha_k d_{ik} + \rho \varepsilon_{it2h}}{\sqrt{1 - \rho^2}} \right) \right],$$

l'égalité correspondant à la limite pour H infini. Il est donc nécessaire de construire un générateur de tirages dans la loi de ε_{it2} conditionnelle à l'observation des $d_{ik}, k=0, \dots, K$. Un moyen d'y parvenir est de disposer dans la seconde équation d'estimateurs convergents de $(\beta_2, \gamma_2, \mu_2)$. On peut alors tirer des ε_{it2} dans une loi normale centrée réduite, et de ne retenir que les tirages vérifiant les observations des $d_{ik}, k=0, \dots, K$, c'est à dire ceux assurant que la variable latente ainsi simulée est dans la bonne tranche. Les tirages ε_{it2h} ainsi retenus sont bien réalisés dans la loi conditionnelle aux $d_{ik}, k=0, \dots, K$; ils peuvent être utilisés pour approximer au moyen d'une simulation l'espérance de y_{it1} décrite ci-dessus :

$$\tilde{m}_i(x_i, d_{ik}, k=0, \dots, K) = \frac{1}{H} \sum_{h=1}^H \left[\Phi \left(\frac{x_{it1} \beta_1 + \sum_{k=0}^K \alpha_k d_{ik} + \rho \varepsilon_{it2h}}{\sqrt{1 - \rho^2}} \right) \right].$$

Il ne reste plus qu'à mettre en regard cette espérance avec les valeurs observées y_{it1} pour obtenir un estimateur des paramètres de la première équation. Un moyen d'y parvenir est d'utiliser la méthode de maximisation du pseudo-maximum de vraisemblance simulé. Comme on l'a signalé, les estimateurs obtenus grâce à cette méthode sont convergents et asymptotiquement normaux pour H infini (GOURIÉROUX, MONFORT [1993]).

En pratique, la méthode est relativement simple à mettre en œuvre ; les calculs s'effectuent en trois étapes :

1- l'estimation des paramètres de la deuxième équation. Une façon simple et robuste d'obtenir des estimateurs convergents des paramètres est d'ajuster le modèle comme un modèle probit polytomique ordonné usuel sur données empilées. On a vu précédemment que cette technique fournit des estimateurs convergents et asymptotiquement normaux des paramètres, et ceci quelque soit la structure temporelle des termes d'erreur ε_{it2} (l'extension au cas du modèle polytomique ordonné est triviale).

2- le tirage des ε_{it2h} , dans la loi normale tronquée, comme décrit précédemment. Une fois déterminés à l'issue de la première étape les estimateurs $(\hat{\beta}_2, \hat{\gamma}_2, \hat{\mu}_2)$, on effectue des tirages indépendants dans une loi normale centrée réduite, en ne retenant que ceux qui correspondent à l'observation de y_{it2} ².

3- l'estimation des paramètres de la première équation. Elle s'effectue avec la mise en œuvre de la méthode du pseudo-maximum de vraisemblance simulé. Une programmation spécifique est cette fois

² Une autre possibilité, préférable, consiste à tirer directement ces résidus dans la loi normale tronquée à l'intervalle correspondant à la fécondité observée. Ceci garantit que tous les tirages « tombent » dans le bon intervalle.

nécessaire, mais celle-ci est relativement légère, et ne fait intervenir que la fonction de répartition de la loi normale à une dimension.

On obtient un estimateur de la matrice de variance-covariance asymptotique des paramètres en recourant à la technique de bootstrap par paires, les termes d'erreur étant I.I.D. par rapport aux individus (EFRON, TIBSHIRANI [1986]). La méthode mise en œuvre consiste à effectuer un tirage avec remise de taille N parmi les N individus (ceux-ci étant indexés par i ; à chaque individu correspondent deux observations pendant T dates). L'estimation en trois étapes sur chaque échantillon bootstrap ainsi constitué fournit des paramètres $\hat{\theta}_b, b=1, \dots, B$ pour les deux équations. Lorsque B tend vers l'infini, la moyenne et la matrice de variance-covariance empiriques des $\hat{\theta}_b, b=1, \dots, B$ sont des estimateurs convergents des paramètres et de leur matrice de variance-covariance.

L'identification du modèle ainsi spécifié n'impose pas a priori que $x_1 \neq x_2$ ni $z_1 \neq z_2$. En raison des restrictions paramétriques introduites par le fait que l'on postule des lois (normales) sur les termes d'erreur, le modèle décrit précédemment est identifiable même si les variables explicatives sont les mêmes dans les deux équations. La situation est d'ailleurs semblable à celle à laquelle on a à faire face lorsque l'on utilise des modèles linéaires : il est nécessaire d'introduire des restrictions identifiantes dans les modèles à équations simultanées si la loi des termes d'erreur n'est pas spécifiée ; dès lors que l'on postule une loi paramétrique sur ces termes, l'identification est assurée (au second ordre) sans qu'il soit nécessaire de recourir à des restrictions identifiantes. Dans des deux cas, on introduit des contraintes au sein du modèle, soit sur les paramètres, soit sur la loi des termes d'erreur. Avec les modèles à variables qualitative, il faut la plupart du temps postuler une loi sur les termes d'erreur, ce qui rend les modèles identifiables sans qu'il soit nécessaire d'introduire des restrictions d'identification. Ceci étant, même si ce n'est pas indispensable, l'identification au premier ordre est souhaitable, et il est préférable de recourir à deux jeux de variables distincts dans les estimations des deux équations.

4. Endogénéité dans un modèle multivarié

Il s'agit ici d'étendre au cas multivarié l'approche adoptée dans le cas du modèle bivarié. Le modèle instrumental est alors polytomique non ordonné, et non plus polytomique ordonné. Il fait donc apparaître autant de variables latentes que d'états dans l'équation de sélection. Les contraintes d'identifiabilité liées à l'invariance des préférences sous-jacentes à toute translation ou homothétie permettent classiquement de faire apparaître un « état de référence » pour lequel la variable latente associée est identiquement nulle. On retient en général le dernier état, si bien que le système instrumental s'écrit :

$$\begin{cases} y_{it2}^{1*} = x_{it2}^1 \beta_2^1 + \varepsilon_{it2}^1 \\ y_{it2}^{2*} = x_{it2}^2 \beta_2^2 + \varepsilon_{it2}^2 \\ \dots \\ y_{it2}^{K*} = 0 \end{cases} .$$

L'équation d'intérêt s'écrit comme précédemment :

$$y_{it1}^* = x_{it1} \beta_1 + \sum_{k=1}^{K-1} \alpha_k d_{ik} + \varepsilon_{it1} .$$

La seule hypothèse sur les lois des termes d'erreur est celle de normalité jointe des $(\varepsilon_{it1}, \varepsilon_{it2}^1, \dots, \varepsilon_{it2}^{K-1})$. Elle permet d'écrire :

$$\varepsilon_{it1} = \sum_{k=1}^{K-1} \rho_k \varepsilon_{it2}^k + w_{it},$$

où w_{it} est une variable aléatoire normale indépendante des ε_{it2}^k . On en déduit :

$$y_{it1}^* = x_{it1} \beta_1 + \sum_{k=1}^{K-1} \alpha_k d_{ik} + \sum_{k=1}^{K-1} \rho_k \varepsilon_{it2}^k + w_{it}.$$

Pour simplifier, on renormalise l'équation d'intérêt en supposant w_{it} de variance unitaire. Pour l'estimation, on utilise la même démarche que dans la partie précédente. Si l'on connaissait les ε_{it2}^k , il deviendrait possible d'estimer l'équation d'intérêt, comme un modèle probit simple. Alors :

$$E(y_{it1} | x_i, d_{ik}, \varepsilon_{it2}^k, k = 1, \dots, K-1) = \Phi(x_{it1} \beta_1 + \sum_{k=1}^{K-1} \alpha_k d_{ik} + \sum_{k=1}^{K-1} \rho_k \varepsilon_{it2}^k).$$

Par conséquent, non conditionnellement à ε_{it2} :

$$E(y_{it1} | x_i, d_{ik}, k = 1, \dots, K-1) = E_{\varepsilon_{it2}^k} \left[\Phi(x_{it1} \beta_1 + \sum_{k=1}^{K-1} \alpha_k d_{ik} + \sum_{k=1}^{K-1} \rho_k \varepsilon_{it2}^k) \right].$$

A nouveau, cette expression fait intervenir l'espérance d'une fonction des variables aléatoires ε_{it2}^k conditionnelle à l'observation des $d_{ik}, k = 1, \dots, K-1$, dont la loi jointe est compliquée. Disposer de tirages $\varepsilon_{it2h}^k, h = 1, \dots, H$ permettrait d'approximer l'espérance par un simulateur :

$$E(y_{it1} | x_i, d_{ik}, k = 0, \dots, K) \approx \frac{1}{H} \sum_{h=1}^H \left[\Phi(x_{it1} \beta_1 + \sum_{k=1}^{K-1} \alpha_k d_{ik} + \sum_{k=1}^{K-1} \rho_k \varepsilon_{it2h}^k) \right],$$

l'égalité correspondant à la limite pour H infini. L'utilisation de la méthode du pseudo-maximum de vraisemblance simulé permet une nouvelle fois d'obtenir des estimateurs convergents et asymptotiquement normaux.

Dans la pratique, la difficulté consiste à réaliser des tirages des ε_{it2}^k conditionnels à l'observation des $d_{ik}, k = 1, \dots, K-1$. La complexité est sérieusement accrue par rapport au cas d'une équation de sélection renvoyant à un modèle polytomique ordonné, puisqu'il s'agit ici de tirer dans une distribution multidimensionnelle tronquée et non plus unidimensionnelle.

Pour conduire l'estimation, on procède à nouveau en trois étapes :

- la première étape consiste à réaliser l'estimation du modèle polytomique non ordonné sur données empilées, comme si on négligeait la corrélation temporelle des termes d'erreur. Cette nouvelle utilisation du pseudo-maximum de vraisemblance permet à nouveau d'obtenir des estimateurs convergents des paramètres β_2^k (GOURIÉROUX, MONFORT, TROGNON [1984]). La plupart des logiciels statistiques disposent de procédures intégrées de calcul d'estimateurs de modèles polytomiques non ordonnés en coupe, cette étape ne pose donc pas de difficulté particulière dès lors que l'on adopte les mêmes hypothèses d'identification que le logiciel utilisé (ce qui peut se révéler une complication en soi).

- la seconde étape, la plus complexe, consiste à simuler les ε_{it2}^k conditionnellement à l'observation des $d_{ik}, k = 1, \dots, K-1$. Une façon d'y parvenir est de recourir à la technique d'échantillonnage de

Gibbs. Il s'agit d'un processus itératif dont l'intuition est la suivante pour le cas qui nous intéresse. On cherche à disposer de tirages dans une loi jointe compliquée, mais les tirages dans chacune des lois marginales conditionnelles à l'observation des tirages dans les autres lois marginales sont relativement simples. En d'autres termes, réaliser des tirages dans la loi conditionnelle de ε_{it}^1 connaissant d_{it}^1 et les autres ε_{it}^k est relativement simple, puisqu'il s'agit de tirer dans une loi normale tronquée. Ce que l'on peut montrer, et qui fait l'intérêt de l'échantillonnage de Gibbs, c'est qu'à force d'itérer des tirages dans ces lois marginales conditionnelles, on finit par obtenir des tirages dans la loi jointe (GEMAN et GEMAN [1984]).

En pratique, on procède de la façon suivante. Soit Ω la matrice de variance des $\varepsilon_{it}^k, k = 1, \dots, K-1$, et c un cycle de rafraîchissement itératif. Au sein du cycle c , lorsque l'on arrive au terme k , on dispose des termes d'erreur $\varepsilon_{it}^{<k(c)}$ correspondant à des valeurs inférieures à k pour le cycle c , et des termes d'erreur $\varepsilon_{it}^{>k(c-1)}$ correspondant à des valeurs supérieures à k pour le cycle $c-1$. Intuitivement, conditionnellement à ces termes d'erreur avant et après k , une partie du terme d'erreur $\varepsilon_{it}^{k(c)}$ à simuler est déjà prédéterminée du fait de la corrélation dès lors que Ω n'est pas diagonale ; seule la partie non corrélée doit être tirée dans une loi normale unidimensionnelle tronquée, ce qui est relativement aisé. Le terme $\varepsilon_{it}^{k(c)}$ est alors la somme de la partie prédéterminée, et du tirage.

Soit par exemple un modèle à trois états, qui fait donc intervenir deux termes d'erreur. On cherche à simuler les résidus pour une observation correspondant au troisième état : l'individu est dans l'état 3 à la date t (on procède de façon analogue si l'individu est dans l'état 1 ou l'état 2, mais il faut effectuer un changement de variables, ce qui rend la présentation plus complexe). Cela signifie que les simulations doivent être telles que :

$$\begin{cases} y_{it}^1 = x_{it}^1 \beta^1 + \varepsilon_{it}^1 \leq 0 \\ y_{it}^2 = x_{it}^2 \beta^2 + \varepsilon_{it}^2 \leq 0 \end{cases}$$

Lors de la première étape, l'estimation du modèle polytomique en coupe a permis de disposer d'estimateurs convergents de (β^1, β^2) et de la matrice de variance :

$$\Omega = \begin{bmatrix} V_1 & C \\ C & V_2 \end{bmatrix}$$

Compte-tenu de la normalité des termes d'erreur, on peut écrire :

$$\begin{cases} \varepsilon_{it}^1 = (C/V_2)\varepsilon_{it}^2 + \sqrt{V_1 - (C^2/V_2)}u_{it}^1 \\ \varepsilon_{it}^2 = (C/V_1)\varepsilon_{it}^1 + \sqrt{V_2 - (C^2/V_1)}u_{it}^2 \end{cases}$$

A l'étape c , connaissant $\varepsilon_{it}^{2(c-1)}$, on tire u_{it}^{1c} dans la loi normale tronquée à l'intervalle :

$$\left] -\infty, -\frac{x_{it}^1 \beta^1 + (C/V_2)\varepsilon_{it}^{2,c-1}}{\sqrt{V_1 - (C^2/V_2)}} \right[$$

On en déduit :

$$\varepsilon_{it}^{1c} = (C/V_2)\varepsilon_{it}^{2,c-1} + \sqrt{V_1 - (C^2/V_2)}u_{it}^{1c}$$

On procède ensuite de même avec u_{it}^{2c} avec un tirage dans l'intervalle :

$$\left[-\infty, -\frac{x_{it2}^2 \beta^2 + (C/V_1) \varepsilon_{it2}^{1,c}}{\sqrt{V_2 - (C^2/V_1)}} \right],$$

afin d'obtenir $\varepsilon_{it2}^{k(c)}$, comme précédemment.

L'algorithme itératif n'est pas très compliqué à programmer. Au bout d'un grand nombre de cycle (en pratique au-delà d'un millier), grâce aux propriétés de l'échantillonnage de Gibbs, les tirages obtenus sont les tirages dans la loi jointe utilisables pour la simulation (HAJIVASSILIOU, MCFADDEN, RUUD [1996]).

- la troisième étape consiste à estimer l'équation d'intérêt. Elle est en tous points analogue à celle décrite pour le modèle polytomique ordonné, et fournit des estimateurs convergents et asymptotiquement normaux des paramètres d'intérêt. La même méthode de bootstrap par paires permet enfin d'obtenir la matrice de variance asymptotique des estimateurs du pseudo-maximum de vraisemblance ainsi calculés.

Conclusion

Les méthodes statistiques destinées à traiter les biais de sélectivité ou d'endogénéité sur données longitudinales quand la variable d'intérêt est quantitative ont donné lieu à une littérature abondante depuis le début des années 1990 (NIJMAN, VERBEEK [1992]). On sait aujourd'hui comment estimer ces modèles, en utilisant diverses méthodes qui diffèrent essentiellement par le caractère plus ou moins paramétrique des hypothèses. Certaines de ces propriétés peuvent être étendues aux modèles dans lesquels la variable d'intérêt est qualitative. Aux procédures reposant sur des hypothèses paramétriques fortes peuvent ainsi être substituées des méthodes d'estimation en plusieurs étapes utilisant les propriétés des pseudo-maximum de vraisemblance, plus économes en matière de spécification, mais aussi beaucoup plus simples à mettre en œuvre.

Références

- BERTSCHEK I. et LECHNER M. (1998) : « Convenient estimators for the panel probit model », *Journal of Econometrics*, 87, p. 329–371.
- BREITUNG J. et LECHNER M. (1996) : « Estimation de modèles non linéaires sur données de panel par la méthode des moments généralisés », *Economie et Prévision*, 126, p. 191-203.
- DUSTMANN C. et ROCHINA-BARRACHINA M. E. (2000) : « Selection correction in panel data models: an application to labour supply and wages », IZA DP, n° 162.
- EFRON B. et TIBSHIRANI R. (1986): « Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy », *Statistical Science*, 1, n°1, p. 54-77.
- GEMAN S., et GEMAN D. (1984) : *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721-724.
- GOURIÉROUX C., MONFORT A. et TROGNON A. (1984) : « Pseudo maximum likelihood methods: Theory », *Econometrica*, 52, p. 681-700.
- GOURIÉROUX C. et MONFORT A. (1993) : « Simulation-based inference : A survey with special reference to panel data models », *Journal of Econometrics*, 59, p. 5-33.
- HAJIVASSILIOU V., MCFADDEN D., et RUUD P. (1996) : « Simulation of multivariate normal rectangle probabilities and their derivatives. Theoretical and computational results », *Journal of Econometrics*, 72, 85-134.
- HECKMAN J. (1974) : « Shadow prices, market prices and labor supply », *Econometrica*, 42, p. 679-693.
- HECKMAN J. (1978): « Dummy endogenous variables in a simultaneous equation system », *Econometrica*, 46, p. 931-959

- KYRIAZIDOU E. (1997) : « Estimation of a panel data sample selection model », *Econometrica*, 65, p. 1335-1364.
- LOLLIVIER S. (2008): « Endogénéité dans les modèles qualitatifs sur données longitudinales », *Annales d'Economie et de Statistiques*, 86, p. 55-76.
- NIJMAN T. et VERBEEK M. (1992) : « Nonresponse in panel data: the impact on estimates of a life cycle consumption function », *Journal of Applied Econometrics*, 7, p. 243- 257.
- ROBIN J.-M. (2000) : « Modèles structurels et variables explicatives endogènes » ; *Méthodologie statistique*, n°2002, INSEE.
- WOOLDRIDGE J. M. (1995) : « Selection corrections for panel data models under conditional mean independence assumptions », *Journal of Econometrics*, 68, p. 115-132.