ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 28 — 1992

Estimation des modèles à deux régimes avec des données de panel

Rachid BOUMAHDI *, Alban THOMAS **

RÉSUMÉ. – Une procédure d'estimation des paramètres d'un modèle avec sélection et régresseurs endogènes est proposée dans le cas de données de panel. La matrice de variance-covariance asymptotique des paramètres est calculée, en adaptant l'approche de LEE, MADDALA et TROST [1980] aux données de panel. Une application du modèle au calcul des rendements de l'éducation est proposée; on montre en particulier la supériorité des matrices d'instruments de BREUSCH-MIZON-SCHMIDT [1989] et AMEMIYA MACURDY [1986] sur celle de HAUSMAN-TAYLOR [1981].

Estimation of Models with Two Regimes and Panel Data

ABSTRACT. – A procedure for estimating parameters of a sample selection model with endogenous regressors is proposed for the case of panel data. The variance-covariance matrix is computed by adapting the LEE, MADDALA and TROST [1980] approach to panel data. An application of the model to the computation of returns to education is proposed; the instrumental matrices of BREUSCH-MIZON-SCHMIDT [1989] and AMEMIYA MACURDY [1986] are shown to be superior to the one of HAUSMAN-TAYLOR [1981].

---

* R. BOUMAHDI: GREMAQ et CEJEE, Université des Sciences Sociales de Toulouse.
** A. THOMAS: INRA, Station d'Économie et Sociologie Rurales de Toulouse.
Les auteurs remercient Jean-Pierre Florens, Peter Schmidt et Quang Vuong pour leurs commentaires et critiques sur une version antérieure de cet article, ainsi que Guy Laroque et deux rapporteurs anonymes.

1 Introduction

Les modèles à variable dépendante tronquée ont connu un développement considérable depuis l'article de TOBIN [1958]. Ce dernier proposait une procédure d'estimation par maximum de vraisemblance d'un modèle à une seule équation. L'idée de base derrière ces modèles est l'existence d'un seuil au-delà duquel la variable dépendante n'est plus observable ou n'a pas de signification économique.

L'estimation des modèles avec variable dépendante tronquée a été étudiée notamment par HECKMAN [1976], [1979], LEE et TROST [1978] pour traiter le biais de sélection dans des modèles à deux régimes et à régresseurs exogènes. AMEMIYA [1978] et, plus récemment, LEE, MADDALA et TROST [1980] – LMT ci-après – ont prolongé la procédure d'estimation en deux étapes, afin de traiter le biais de sélection dans des modèles à équations simultanées.

Ce courant n'a cependant pas, à notre connaissance, encore intégré le domaine de l'économétrie des données de panel, qui s'est limitée à l'estimation de modèles à un seul régime. L'approche de BALESTRA et NERLOVE [1966] a été enrichie par l'introduction de régresseurs exogènes (MADDALA [1971] et NERLOVE [1971]) et endogènes (MUNDLAK [1978], HAUSMAN et TAYLOR [1981] – HT ci-après –, AMEMIYA et MACURDY [1986] – AM ci-après – et BREUSCH, MIZON et SCHMIDT [1989] – BMS ci-après).

La présence de régresseurs endogènes dans un modèle avec données de panel nécessite un traitement particulier en raison de la corrélation entre les effets individuels et certaines variables explicatives. HT, AM et BMS proposent à cet égard des matrices d'instruments pour l'estimation de la forme structurelle du modèle.

Si l'on considère l'extension de ce modèle au cas de deux régimes, une seconde source d'endogénéité peut alors apparaître. En effet, les termes de perturbation de l'équation de sélection utilisée pour affecter les individus à l'un des deux régimes sont susceptibles d'être corrélés avec les effets individuels. Les procédures de HT, AM et BMS restent utilisables pour traiter la première source d'endogénéité, mais il est nécessaire de corriger également le biais de sélection.

Dans cet article, une procédure d'estimation des paramètres d'un modèle avec sélection et régresseurs endogènes est proposée dans le cas de données de panel. L'hétéroscédasticité provenant de la structure du modèle à erreurs composées est tout d'abord éliminée par la diagonalisation de la matrice de variance-covariance. Le biais de sélection est ensuite corrigé par la méthode classique de l'introduction dans la forme structurelle de l'espérance conditionnelle des erreurs. L'endogénéité de certaines variables explicatives nécessite enfin le recours à des matrices d'instruments (méthode de HT, AM et BMS). Les paramètres de la forme structurelle sont présentés et la matrice de variance-covariance asymptotique des paramètres est calculée, en adaptant l'approche de LMT aux données de panel.

La section 2 présente le modèle à deux régimes avec sélection et introduit les notations utilisées. La méthode d'estimation utilisée pour la forme structurelle de notre modèle, ainsi que l'expression de la matrice de variance-covariance asymptotique des paramètres figure dans la section 3. A titre d'illustration, la section 4 présente une application du modèle au calcul des rendements de l'éducation. Une équation de salaire est estimée en utilisant des données d'enquête individuelles et en discriminant les individus en fonction de leur passage ou non dans l'enseignement supérieur. Le calcul des estimateurs de la variance des erreurs est fourni en annexe.

2 Le modèle

On considère une population d'individus de taille N, constituée de deux sous-ensembles disjoints N_1 et N_2, de sorte que $N_1 + N_2 = N$. Le comportement d'un individu peut être représenté par l'un des deux régimes I_j, $j = 1, 2$, selon qu'il appartient à la sous-population N_1 ou N_2. Cette appartenance est observée *ex ante*. L'affectation de l'individu à l'un de ces régimes est déterminée par une équation de sélection de la forme :

$$y_i^* = \omega_i \lambda + v_i \tag{1}$$

où y_i^* est une variable latente non observable et ω_i un vecteur de variables exogènes. On suppose que l'individu i appartient au premier régime si $y_i^* > 0$. On a par conséquent

$$\Pr[i \in I_1] = \Pr[y_i^* > 0] = \Pr[\omega_i \lambda > -v_i] \tag{2}$$

et

$$\Pr[i \in I_2] = \Pr[y_i^* \leqslant 0] = \Pr[\omega_i \lambda \leqslant -v_i] \tag{3}$$

L'équation de sélection ne dépend pas du temps, ce qui signifie que l'individu est affecté à l'un des deux régimes de façon définitive. La variable dépendante, notée Y_{jit} (observation à la période t pour l'individu i appartenant au régime j) est expliquée par des variables X_{jit} et Z_{ji}, respectivement variantes et non variantes dans le temps, et contenant des endogènes.

Le modèle à deux régimes considéré s'écrit :

$$Y_{it} = Y_{1\,it} = \beta_1' X_{1\,it} + \gamma_1' Z_{1\,i} + \alpha_{1\,i} + \varepsilon_{1\,it} \quad \text{si} \quad \omega_i \lambda > -v_i \tag{4}$$

$$Y_{it} = Y_{2\,it} = \beta_2' X_{2\,it} + \gamma_2' Z_{2\,i} + \alpha_{2\,i} + \varepsilon_{2\,it} \quad \text{si} \quad \omega_i \lambda \leqslant -v_i \tag{5}$$

où $i = 1, \ldots, N_1, N_1 + 1, \ldots, N$; $t = 1, \ldots, T$.

Dans la suite de l'article, on utilisera la forme plus compacte suivante

$$Y_1 = X_1 \beta_1 + Z_1 \gamma_1 + \alpha_1 + \varepsilon_1, \qquad i \in I_1 \tag{6}$$

$$Y_2 = X_2 \beta_2 + Z_2 \gamma_2 + \alpha_2 + \varepsilon_2, \qquad i \in I_2 \tag{7}$$

où $Y_j = [Y_{j11}, \ldots, Y_{j1T}, Y_{j21}, \ldots, Y_{j2T}, \ldots, Y_{jN1}, \ldots, Y_{jN_jT}]'$ est un vecteur $TN_j \times 1$; X_j est $TN_j \times K_j$, Z_j est $TN_j \times G_j$. α_j est le vecteur $TN_j \times 1$ des effets individuels et ε_j le vecteur $TN_j \times 1$ des termes d'erreur variant dans le temps. Les termes z_{ji} et α_{ji} ne variant pas dans le temps, les vecteurs Z_j et α_j s'écrivent

$$Z_j = \text{vec}[z_1 \iota_T, z_2 \iota_T, \ldots, z_{N_j} \iota_T]$$

$$\alpha_j = \text{vec}[\alpha_1 \iota_T, \alpha_2 \iota_T, \ldots, \alpha_{N_j} \iota_T]$$

où ι_T est un vecteur de dimension $T \times 1$ dont les éléments sont égaux à 1.

Les coefficients β_j, γ_j sont des vecteurs de dimension $(K_j \times 1)$, $(G_j \times 1)$ associés respectivement aux variables variantes et non variantes dans le temps. Le système [6]-[7] décrit des données de panel avec discrimination des individus en fonction de leurs caractéristiques individuelles. On fera les hypothèses classiques des modèles à erreurs composés en ce qui concerne la distribution des erreurs: les effets individuels α_j, $j = 1, 2$ sont i.i.d. de loi $N(0, \sigma^2_{\alpha_j})$, supposés indépendants des ε_j et corrélés avec v, le vecteur $N \times 1$ des perturbations provenant de l'équation de sélection.

De plus, les termes d'erreur ε_j sont i.i.d. $N(0, \sigma^2_{\varepsilon_j})$ et indépendants de v. Enfin, on suppose que la covariance entre α_1 et α_2 d'une part, ε_1 et ε_2 d'autre part est nulle.

Le caractère endogène d'une partie de X_j et des Z_j est pris en compte de la façon suivante. Suivant HT, on partitionne les matrices de variables explicatives en

$$X_j = [\dot{X}_j : \ddot{X}_j] \quad \text{et} \quad Z_j = [\dot{Z}_j : \ddot{Z}_j], \qquad j = 1, 2$$

de dimension $[TN_j \times \dot{K}_j : TN_j \times \ddot{K}_j]$ et $[TN_j \times \dot{G}_j : TN_j \times \ddot{G}_j]$, où $\dot{K}_j + \ddot{K}_j = K_j$ et $\dot{G}_j + \ddot{G}_j = G_j$. Les variables contenues dans les sous-matrices $\ddot{X}_j$ et $\ddot{Z}_j$ sont endogènes, celles dans $\dot{X}_j$ et $\dot{Z}_j$ sont exogènes. La seule source d'endogénéité vient alors de l'hypothèse de corrélation entre les effets individuels α_j et $\ddot{X}_j$, $\ddot{Z}_j$.

Pour les divers projecteurs, on utilisera les notations suivantes. Soit M_j la matrice des variables dichotomiques individuelles de dimension $TN_j \times N_j$:

$$M_j = I_{N_j} \otimes \iota_T, \qquad j = 1, 2.$$

On définit la matrice de projection P_{M_j} par

$$P_{M_j} = M_j (M_j' M_j)^{-1} M_j' = I_{N_j} \otimes \frac{1}{T} \iota_T \iota_T', \qquad j = 1, 2 \tag{8}$$

(projection sur le sous-espace engendré par les colonnes de M_j) et

$$Q_{M_j} = I_{TN_j} - P_{M_j}, \qquad j = 1, 2 \tag{9}$$

(projection sur le sous-espace orthogonal à celui engendré par les colonnes de M_j).

Par construction, P_{M_j} transforme un vecteur d'observations de dimension $TN_j \times 1$ tel que Y_j en un vecteur des moyennes individuelles, alors que Q_{M_j} le transforme en un vecteur d'écarts à la moyenne individuelle. Avec ces notations, la matrice de variance-covariance (de dimension $TN_j \times TN_j$) du terme d'erreur s'écrit

$$(10) \quad E(\alpha_j+\varepsilon_j)(\alpha_j+\varepsilon_j)'=\Omega_j=\sigma^2_{\varepsilon_j} Q_{M_j}+(\sigma^2_{\varepsilon_j}+T\,\sigma^2_{\alpha_j})P_{M_j}, \qquad j=1, 2$$

3 Méthode d'estimation

Trois problèmes doivent être traités, de façon séquentielle, afin de fournir des estimateurs non biaisés et convergents. Tout d'abord, le modèle à erreurs composées implique une structure hétéroscédastique que l'on corrige en diagonalisant les matrices de variance-covariance. Ensuite, la corrélation entre les effets individuels et les erreurs de l'équation de sélection introduit un biais de sélection. Ce problème est résolu en utilisant une méthode de type probit, consistant à calculer l'espérance du terme d'erreur structurel. Enfin, l'existence de régresseurs endogènes dans le modèle doit être prise en compte dans le calcul des estimateurs; il est donc nécessaire d'utiliser la méthode des variables instrumentales.

Considérons tout d'abord le cas d'un modèle à un seul régime [équation [6]) avec régresseurs endogènes. Il n'y a pas de problème de biais de sélection et HT ont proposé un estimateur efficace des deux paramètres β_1 et γ_1. Cet estimateur exploite l'hypothèse selon laquelle $\dot{X}_1$ et $\dot{Z}_1$ sont indépendants de l'effet individuel α_1. AM ont suggéré un estimateur plus efficace que celui de HT, fondé sur des hypothèses d'exogénéité très fortes. En fait, la différence entre les deux estimateurs réside dans le traitement des variables variantes dans le temps non corrélées avec les effets individuels.

A cet égard, HT supposent que la moyenne des $\dot{X}_1$ est indépendante de α_1; ils utilisent chaque variable dans $\dot{X}_1$ comme deux instruments: comme moyenne et comme écart à la moyenne. En revanche, AM imposent une hypothèse d'exogénéité selon laquelle $\dot{X}_1$ est indépendant de α_1 à chaque période t, $t=1, \ldots, T$. Les auteurs utilisent $\dot{X}_1$ comme $(T+1)$ instruments:

comme écart à la moyenne et séparément pour chaque période. Plus récemment, BMS ont prolongé l'analyse de AM pour dériver un estimateur plus efficace que les deux précédents. [1]

Le passage à deux régimes introduit une corrélation supplémentaire, celle entre α_1 et v. La méthode de HT est impraticable dans ce cas, car elle fournirait des estimateurs biaisés. Pour simplifier la présentation, on s'intéresse uniquement à l'estimation du premier régime. [2]

On a

$$\begin{aligned} (11) \qquad & E[\alpha_{1\,i}+\varepsilon_{1\,it} \mid X_{1\,it}, Z_{1\,i}, \omega_i \lambda > -v_i] \\ & = E[\alpha_{1\,i} \mid X_{1\,it}, Z_{1\,i}, \omega_i \lambda > -v_i] \\ & + E[\varepsilon_{1\,it} \mid X_{1\,it}, Z_{1\,i}, \omega_i \lambda > -v_i] \\ & = E[\alpha_{1\,i} \mid X_{1\,it}, Z_{1\,i}, \omega_i \lambda > -v_i] \neq 0 \end{aligned}$$

Dans le contexte d'un modèle à deux régimes (coupe instantanée), LMT ont proposé l'estimation de deux systèmes d'équations simultanées (pour une coupe instantanée) en présence de biais de sélection. Leur procédure consiste à estimer les formes réduite et structurelle de chaque système en tenant compte du biais de sélection et à dériver une matrice asymptotique de variance-covariance des paramètres pour les deux formes réduite et structurelle. Pour estimer le système d'équations [6] et [7], nous pouvons donc procéder de la même manière que LMT. Notre approche va consister à utiliser conjointement les deux méthodes de HT et LMT.

Une première étape consiste à corriger l'hétéroscédasticité en diagonalisant la matrice Ω_1. Soit

$$\Omega_1^{-1/2} = Q_{M_1} + \theta_1 P_{M_1}$$

où

$$\theta_1 = \frac{\sigma_{\varepsilon_1}}{(\sigma_{\varepsilon_1}^2 + T\,\sigma_{\alpha_1}^2)^{1/2}}$$

La matrice $\Omega_1^{-1/2}$ transforme Ω_1 en matrice diagonale; en d'autres termes,

$$\Omega_1^{-1/2}\,\Omega_1\,\Omega_1^{-1/2} = \sigma_{\varepsilon_1}^2\, I_{TN_1}$$

En fait, on utilise un estimateur convergent $\hat{\Omega}_1^{-1/2}$ de $\Omega_1^{-1/2}$, basé sur un estimateur $\hat{\sigma}_{\varepsilon_1}^2$ obtenu par la méthode de HT, et un estimateur $\hat{\sigma}_{\alpha_1}^2$ calculé en tenant compte du biais de sélection (voir l'annexe).

1. HT et AM utilisent comme instrument l'écart à la moyenne des variables variantes dans le temps, alors que BMS utilisent en plus les (T − 1) instruments linéairement indépendants de ces écarts. Pour une revue plus détaillée, voir BMS.
2. La procédure est identique pour le second régime.

En multipliant [6] par $\Omega^{-1/2}$, on obtient:

$$\Omega_1^{-1/2} Y_1 = \Omega_1^{-1/2} X_1 \beta_1 + \Omega_1^{-1/2} Z_1 \gamma_1 + \Omega_1^{-1/2} \alpha_1 + \Omega_1^{-1/2} \varepsilon_1 \quad \text{si} \quad i \in I_1 \tag{12}$$

Considérons à présent le problème du biais de sélection. L'équation structurelle corrigée de l'hétéroscédasticité se réécrit:

$$\Omega_1^{-1/2} Y_1 = \Omega_1^{-1/2} X_1 \beta_1 + \Omega_1^{-1/2} Z_1 \gamma_1 + \Omega_1^{-1/2} (\alpha_1 + \varepsilon_1) \tag{13}$$

ou de façon équivalente:

$$\Omega_1^{-1/2} Y_1 = \Omega_1^{-1/2} \dot{X}_1 \dot{\beta}_1 + \Omega_1^{-1/2} \ddot{X}_1 \ddot{\beta}_1 + \Omega_1^{-1/2} \dot{Z}_1 \dot{\gamma}_1 + \Omega_1^{-1/2} \ddot{Z}_1 \ddot{\gamma}_1 + \Omega_1^{-1/2} (\alpha_1 + \varepsilon_1) \tag{14}$$

où

$$\mathrm{Var}[\Omega^{-1/2} (\alpha_1 + \varepsilon_1)] = \sigma_{\varepsilon_1}^2 I_{TN_1}$$

et

$$\beta_1 = \begin{pmatrix} \dot{\beta}_1 \\ \ddot{\beta}_1 \end{pmatrix}, \qquad \gamma_1 = \begin{pmatrix} \dot{\gamma}_1 \\ \ddot{\gamma}_1 \end{pmatrix}$$

L'espérance conditionnelle de Y_1^* est

$$E[Y_1^* \mid i \in I_1] = \dot{X}_1^* \dot{\beta}_1 + \ddot{X}_1^* \ddot{\beta}_1 + \dot{Z}_1^* \dot{\gamma}_1 + \ddot{Z}_1^* \ddot{\gamma}_1 + E[\Omega_1^{-1/2} \alpha_1 + \Omega_1^{-1/2} \varepsilon_1 \mid i \in I_1] \tag{15}$$

où

$$Y_1^* = \Omega_1^{-1/2} Y_1; \qquad X_1^* = [\dot{X}_1^* : \ddot{X}_1^*] = \Omega_1^{-1/2} X_1$$
$$Z_1^* = [\dot{Z}_1^* : \ddot{Z}_1^*] = \Omega_1^{-1/2} Z_1$$

On a

$$E[\Omega_1^{-1/2} \varepsilon_1 \mid i \in I_1] = 0 \quad \text{et} \quad E[\Omega_1^{-1/2} \alpha_1 \mid i \in I_1] = E[\theta_1 \alpha_1 \mid i \in I_1] = -\sigma_{v\alpha_1^*} \frac{\phi}{\Phi}$$

où $\sigma_{v\alpha_1^*} = \mathrm{Cov}(v, \theta_1 \alpha_1) = \mathrm{Cov}(v, \alpha_1^*)$ et $\frac{\phi}{\Phi}$ est le ratio de Mill, défini comme un vecteur $TN_1 \times 1$ dont l'élément générique représente le rapport entre la densité et la fonction de répartition de la loi normale centrée réduite au point $-v_i$, lequel provient de l'estimation préalable de l'équation de sélection.

λ est tout d'abord calculé par maximum de vraisemblance. En remplaçant λ par son estimation $\hat{\lambda}$ on obtient

$$Y_1^* = \dot{X}_1^* \dot{\beta}_1 + \ddot{X}_1^* \ddot{\beta}_1 + \dot{Z}_1^* \dot{\gamma}_1 + \ddot{Z}_1^* \ddot{\gamma}_1 - \sigma_{v\alpha_1^*} \frac{\hat{\phi}}{\hat{\Phi}} + \tilde{u}_1 \tag{16}$$

où $\tilde{u}_1$ est un vecteur $TN_1 \times 1$ de perturbations tel que

$$E[\tilde{u}_1 \mid i \in I_1] = 0$$

On pose

$$H = \left[X_1^*, Z_1^*, -\frac{\hat{\phi}}{\hat{\Phi}} \right], \qquad \Theta = \begin{pmatrix} \beta_1 \\ \gamma_1 \\ \sigma_{\upsilon\alpha_1^*} \end{pmatrix}$$

L'estimateur de la variable instrumentale ou des doubles moindres carrés est donné par

(17) $$\hat{\Theta} = (H' P_{A_1} H)^{-1} H P_{A_1} Y_1^*$$

avec $P_{A_1} = A_1 (A_1' A_1)^{-1} A_1'$, où A_1 est la matrice des instruments. Pour HT par exemple, on a $A_1 = (Q_{M_1}, P_{M_1} \dot{X}_1, \dot{Z}_1)$. L'estimateur $\hat{\Theta}$ est un estimateur convergent. La matrice de variance-covariance asymptotique est

(18) $$\begin{aligned} \operatorname{Var}(\hat{\Theta}) = (\sigma_{\alpha_1}^2 + \sigma_{\varepsilon_1}^2)(H' P_{A_1} H)^{-1} - \sigma_{\upsilon\alpha_1^*} (H' P_{A_1} H)^{-1} H' P_{A_1} \\ \times [B - BW_1 (W' \Lambda W)^{-1} W_1' B]^{-1} P_{A_1} H (H' P_{A_1} H)^{-1} \end{aligned}$$

où W, W_1, B et Λ sont définies comme suit :

$$W = (w_{11} \ldots w_{1T}, \ldots, w_{N_1 1} \ldots w_{N_1 T}, \ldots, w_{N1} \ldots w_{NT}) = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix}$$

est la matrice de dimension $T(N_1 + N_2) \times 1$ des variables exogènes dans l'équation de sélection;

$$B = \begin{bmatrix} w_1 \lambda \dfrac{\phi_1}{\Phi_1} + \left(\dfrac{\phi_1}{\Phi_1}\right)^2 \ldots & 0 \\ 0 \quad \ldots & 0 \\ 0 \quad \ldots & w_{N_1} \lambda \dfrac{\phi_{N_1}}{\Phi_{N_1}} + \left(\dfrac{\phi_{N_1}}{\Phi_{N_1}}\right)^2 \end{bmatrix} \otimes I_T$$

de dimension $TN_1 \times TN_1$.

$$\Lambda = \begin{bmatrix} \dfrac{\phi_1^2}{\Phi_1 (1 - \Phi_1)} \ldots & 0 \\ 0 \quad \ldots & 0 \\ 0 \quad \ldots & \dfrac{\phi_N^2}{\Phi_N (1 - \Phi_N)} \end{bmatrix} \otimes I_T$$

est la matrice de dimension $TN \times TN$ dont le terme générique représente la contribution de chaque observation i, $i = 1, \ldots, N$.

LMT montrent que cette matrice de variance-covariance est la seule pertinente pour les modèles à équations simultanées avec sélection. Par rapport à l'approche de LMT, les dimensions des matrices H, W, B et Λ doivent bien entendu être adaptées au cas des données de panel.

4 Application au calcul des rendements de l'éducation

Dans cette section, on propose une application du modèle décrit dans la section 2 au calcul des rendements de l'éducation avec biais de sélection (voir WILLIS et ROSEN [1979], BOUMAHDI *et al.* (1990]). Ce modèle permet de mesurer l'avantage relatif en termes de salaire d'un passage dans un cycle d'études supérieur. L'un des points les plus intéressants est de pouvoir comparer les performances relatives des méthodes de HT, AM et BMS, qui diffèrent dans la construction de la matrice d'instruments. Cette comparaison est bien sûr basée sur la matrice de variance-covariance des paramètres estimés.

On considère deux niveaux d'éducation E_1 et E_2 (enseignement supérieur et secondaire respectivement); Y_1 et Y_2 indiquent les logarithmes des salaires associés à E_1 et E_2 respectivement. La condition $\omega_i \lambda > -v_i$ désigne ici le passage du régime E_2 au régime E_1 de l'individu *i*. La matrice $W = [\omega'_1, \ldots, \omega'_N]$ comprend des variables d'aptitude et d'opportunité :

- EP2 à EP5 : variables indicatrices liées au niveau d'éducation du père. EP2 = 1 pour le certificat d'études, EP3 = 1 pour le BEPC, EP4 = 1 pour le Baccalauréat et EP5 = 1 pour un diplôme supérieur;
- EM2 à EM5 : variables indicatrices liées au niveau d'éducation de la mère. Les définitions sont les mêmes que pour EP2 à EP5;
- FS : nombre de frères et sœurs de l'individu l'année où il termine ses études;
- SEX : variable indicatrice égale à 1 pour les hommes et 0 pour les femmes;
- PRIM : nombre d'années d'études passées dans l'enseignement primaire;
- NRS : nombre de redoublements durant l'enseignement secondaire.

La disponibilité des données a été déterminante dans le choix de la formulation de l'équation de sélection. En effet, en absence d'une mesure spécifique de l'aptitude, on a admis que cette dernière pouvait être captée par les deux variables : nombre d'années d'études passées dans l'enseignement primaire, PRIM, et le nombre de redoublements dans l'enseignement secondaire, NRS.[3] La deuxième catégorie de variables jouant un rôle déterminant sur la probabilité du passage d'un individu à l'enseignement supérieur est représentée par les variables d'opportunité mesurées notamment par l'enseignement général et supérieur des parents, EP_j, EM_j,

3. Pour une discussion détaillée, voir BOUMAHDI [1991 *a*].

$j=1, \ldots, 5$ et le nombre de frères et sœurs de l'individu l'année où il termine ses études, FS.

Le logarithme du salaire, SAL, est expliqué par les variables suivantes :

• exogènes et variant dans le temps ($\dot{X}_1$) : région (RG = 1 si l'individu réside dans la région parisienne, 0 sinon) et le statut marital (MS = 1 si l'individu est marié, 0 sinon);

• exogènes et invariant dans le temps ($\dot{Z}_1$) : sexe (SEX = 1 pour un homme, 0 pour une femme);

• endogènes et variant dans le temps ($\ddot{X}_1$) : expérience (EXP, le nombre d'années d'activité) et l'expérience au carré (EXP^2);

• endogènes et invariant dans le temps ($\ddot{Z}_1$) : nombre d'années passées dans le système éducatif (EDUC). [4]

L'effet individuel α_j représente ici l'aptitude, qui est corrélée avec l'éducation et l'expérience d'une part, avec v d'autre part (voir CHAMBERLAIN et GRILICHES [1975], GRILICHES [1977], BOUMAHDI et PLASSARD [1992]). Remarquons que la corrélation entre α_j (variable non observable) et v apparaît très justifiée. En effet, le passage du niveau d'éducation E_2 au niveau E_1 peut être expliqué par la matrice W (variables d'aptitude et d'opportunité) mais aussi par l'effet individuel α_1.

Il existe dans différents pays des ensembles de données très complets permettant d'estimer notre modèle. Par exemple aux États-Unis, l'enquête PSID (voir HAUSMAN-TAYLOR [1981]) permet des études économétriques sur des individus identiques suivis pendant plus de vingt ans. Malheureusement, en France, l'on ne dispose pas de données aussi complètes. L'une des bases de données les plus exhaustives concernant les qualifications professionnelles en France est l'enquête FQP 85 (Formation et Qualification Professionnelle) de l'INSEE, qui est réalisée avec des individus différents, tous les sept ans. Pour appliquer les méthodes préconisées ci-dessus, nous avons donc construit trois coupes [1986, 1987 et 1988] à partir d'une première coupe [1985] comportant des données réelles. Les valeurs des variables variant dans le temps ont été simulées pour chaque individu en utilisant des données statistiques agrégées. L'on disposera donc de quatre coupes instantanées portant sur les mêmes individus.

Les variables devant être simulées sont SAL, RG et MS (toutes les autres variables sont soit invariantes dans le temps, soit déterministes). Les valeurs du logarithme du salaire, SAL, de 1986 à 1988 sont calculées en utilisant les taux de croissance annuels du salaire en fonction du sexe et de la catégorie socio-professionnelle de l'individu (INSEE [1990]).

Pour les variables RG et MS, la procédure est la suivante. A partir de statistiques sur la population active en France, on calcule tout d'abord les probabilités de changement de statut (par exemple un mariage ou un

4. Remarquer que la condition d'identification pour HT : $\dot{K}_1 \geqslant \ddot{G}_1$ est satisfaite : $\dot{K}_1 = 2$ (RG, MS) et $\ddot{G}_1 = 1$ (EDUC). Elle l'est également pour AM et BMS.

déménagement vers une autre région) en utilisant les fréquences correspondantes dans la population active totale. De telles fréquences varient bien entendu selon, en particulier, le sexe et l'âge. Pour chaque individu i et chaque statut susceptible de varier (RG et MS), nous tirons ensuite une variable aléatoire à partir d'une distribution uniforme U [0; 1]. Le statut de l'individu d'une année sur l'autre est changé si la valeur tirée est inférieure à la probabilité théorique correspondant à ce même statut, étant donné l'âge, le sexe et la catégorie socio-professionnelle de l'individu.

Les probabilités de changement de région (région parisienne vers province et vice-versa, variable RG) sont calculées d'après les données de 1982 sur la mobilité régionale (INSEE [1982]).

Nous supposons que le statut marital, MS, n'est modifié qu'en cas de (re)mariage ou de divorce; de plus, si un individu divorce, on suppose qu'il ne se remarie pas durant le reste de la période. Les taux de mariage et de divorce par sexe et par âge sont utilisés pour calculer les probabilités individuelles correspondantes (INSEE [1990]):

$$\mathrm{Prob}[\mathrm{MS}_{it}=1 \mid \mathrm{MS}_{i,t-1}=0] \qquad \text{et} \qquad \mathrm{Prob}[\mathrm{MS}_{it}=0 \mid \mathrm{MS}_{i,t-1}=1]$$

Des précautions doivent être prises pour empêcher des modifications trop fréquentes du statut des individus. Puisque l'on dispose de quatre périodes, il est raisonnable de supposer qu'une caractéristique particulière ne peut changer qu'une seule fois de 1985 à 1988. Une telle procédure est bien sûr criticable mais elle est néanmoins utilisée ici pour donner au lecteur une idée des méthodes et de leur applicabilité, plutôt que des résultats numériques précis.

A cet égard, l'on peut s'interroger sur l'effet de la procédure de construction des données simulées sur les résultats d'estimation. Le fait de simuler des valeurs pour la variable dépendante SAL, par un argument d'erreur sur les variables, entraîne une sous-estimation des écart-types. En effet, l'erreur de « mesure » due à la simulation vient s'ajouter au terme d'erreur structurel de notre modèle, ce qui augmente la variance de la perturbation totale. Comme cet effet n'est pas pris en compte, les écart-types des paramètres sont sous-estimés. Ce problème ne nous semble pas trop important, dans la mesure où cette situation est valable pour les trois méthodes décrites (HT, AM et BMS), et que notre objectif est de comparer ces trois dernières. En ce qui concerne les régresseurs, les variables RG et MS (Région et État Matrimonial) ne sont pas corrélées avec l'effet individuel, puisqu'elles sont exogènes dans notre modèle. Leur simulation ne devrait donc pas avoir un effet trop néfaste sur la comparaison entre les trois instruments.

Notons que l'objectif majeur de cette application empirique, sur le calcul des rendements de l'éducation, est de montrer le gain d'efficacité potentiel lié à l'utilisation des différents instruments pour traiter la covariance entre les effets individuels et certaines variables explicatives.

L'échantillon utilisé comprend des actifs, hommes et femmes, âgés de 35 à 40 ans. Il y a 177 individus dans le premier régime (enseignement supérieur) et 325 dans le second, de sorte que NT = 2008. Dans ce qui suit, on considérera uniquement l'estimation du premier régime, la procédure étant similaire pour le second.

Les statistiques descriptives des variables composant l'échantillon du premier régime figurent dans le tableau 1.

TABLEAU 1

Statistiques descriptives des variables de l'échantillon (études supérieures)

Variable et année	Moyenne	Écart-type	Min.	Max
MS 1985	0,0508	0,2203	0	1
MS 1986	0,0791	0,2706	0	1
MS 1987	0,0960	0,2954	0	1
MS 1988	0,1469	0,3550	0	1
RG 1985	0,2994	0,4593	0	1
RG 1986	0,2768	0,4487	0	1
RG 1987	0,2485	0,4334	0	1
RG 1988	0,2655	0,4428	0	1
SAL 1985	135 870,7	58 312,13	35 000	364 788
SAL 1986	141 911,5	62 208,36	37 058,8	384 122,1
SAL 1987	146 759	65 695,82	38 162,5	400 560,6
SAL 1988	151 301	68 802,08	39 481,6	415 417,4

Le tableau 2 donne les résultats d'estimation de l'équation de sélection. A partir de ces estimations, le rapport $\frac{\phi}{\Phi}$ est aisément calculé; il sera utilisé plus tard comme régresseur dans l'équation structurelle. La matrice $W = [\omega_1, \ldots, \omega_N]$ sera également nécessaire, pour le calcul de la matrice de variance-covariance.

La variable PRIM a un cœfficient négatif et significatif; par conséquent, une scolarité très longue dans l'enseignement primaire diminue, toutes choses égales par ailleurs, la probabilité de passage à l'enseignement supérieur. Les variables d'opportunité donnent elle aussi des résultats cohérents avec ce que l'on était en droit d'attendre. L'éducation des parents a un effet positif sur la scolarité supérieure de l'enfant, un nombre élevé de frères et de sœurs implique une offre de financement plus faible par enfant et par voie de conséquence – comme le financement de l'investissement éducatif revient en grande partie aux parents – diminue les chances de poursuivre des études post-secondaires.

L'estimation de l'équation de sélection permet de calculer le ratio de Mill utilisé ensuite comme variable explicative dans l'équation structurelle du premier régime [équation (16)].

L'étape suivante consiste à calculer les variances $\sigma^2_{\alpha_1}$ et $\sigma^2_{\varepsilon_1}$ des termes d'erreur (voir l'annexe pour plus de détails). La matrice de variance-covariance $\Omega_1^{-1/2}$ peut dès lors être construite. Nous estimons ensuite les paramètres β_1 et γ_1 pour le premier régime, dans l'équation corrigée de l'hétéroscédasticité.

TABLEAU 2

Estimation des paramètres de l'équation de sélection

Variable	Estimation	Écart-type
Constante	1,171 67	0,482 0
EP2	−0,014 8	0,185 6
EP3	0,505 8	0,260 1
EP4	0,316 6	0,342 4
EP5	0,836 5	0,295 5
EM2	0,402 3	0,193 8
EM3	0,879 3	0,241 8
EM4	1,262 1	0,343 4
EM5	1,107 2	0,401 0
FS	−0,070 2	0,037 2
SEX	0,173 7	0,140 8
PRIM	−0,393 2	0,073 8
NRS	−0,011 2	0,093 2

Les matrices d'instruments de HT, AM et BMS sont utilisées successivement dans les estimations. Pour chaque méthode, les écart-types des paramètres estimés donnent une indication sur la performance relative de chacune.

Les matrices d'instruments sont calculées comme suit :

$$A_{HT} = (Q_{M_1}, P_{M_1}\dot{X}_1, \dot{Z}_1)$$

$$A_{AM} = (Q_{M_1}, P_{M_1}\dot{X}_1, Q_{M_1}\dot{X}_1^0, \dot{Z}_1)$$

$$A_{BMS} = (Q_{M_1}, P_{M_1}\dot{X}_1, Q_{M_1}X_1^0, \dot{Z}_1).$$ [5]

Rappelons que la procédure d'estimation proposée ici est basée sur l'hypothèse de corrélation entre $(\ddot{X}_1, \ddot{Z}_1)$ et l'effet individuel α_1. HAUSMAN [1978, p. 1263] propose un test d'exogénéité afin de pouvoir tester l'hypothèse nulle $H_0: E[\alpha_1 | X_1, Z_1] = 0$ contre l'hypothèse alternative $H_1: E[\alpha_1 | X_1, Z_1] \neq 0$. Ce dernier consiste à comparer l'estimateur within-groups (intra-groupes) à celui des moindres carrés généralisés. Sous l'hypothèse nulle, la valeur de la statistique de test, dans le cas présent, est 177.13. On peut donc rejeter H_0 avec un risque pratiquement nul. [6]

Le tableau 3 donne l'estimation de la fonction de gains [régime supérieur, équation [(16)], la formule (18) est utilisée pour calculer la matrice de variance-covariance des paramètres. Les estimations des moindres carrés généralisés sont également présentés.

Le logarithme du salaire apparaît comme une fonction croissante et concave de l'expérience. Le coefficient de la variable EDUC, qui s'interprète

5. Pour la définition de la matrice $\dot{X}_1^0$, voir BMS.

6. Dans le calcul de la statistique de test de Hausman, la matrice d'instruments de BMS est utilisée pour estimer $\sigma^2_{\alpha_1}$, cette dernière intervenant dans le calcul de la matrice $\hat{\Omega}_1^{-1/2}$.

TABLEAU 3

Estimation des paramètres structurels et écart-types (Méthodes des moindres carrés généralisés (GLS), HT, AM et BMS)*

Variable	GLS	HT	AM	BMS
Constante	0,154 10	−0,332 13	−0,263 15	−0,268 32
	(0,006 1)	(4,684)	(0,725)	(0,602)
RG	−0,025 05	−0,025 40	−0,021 33	−0,022 20
	(0,006 9)	(0,464)	(0,111)	(0,111)
MS	0,001 32	0,001 32	0,003 96	0,004 30
	(0,006 0)	(0,391)	(0,149)	(0,148)
EXP	0,037 57	0,037 61	0,038 10	0,036 64
	(0,004 7)	(0,311)	(0,149)	(0,149)
EXP^2	−0,000 20	−0,000 20	−0,000 24	−0,000 21
	(0,000 1)	(0,010)	(0,004)	(0,004)
SEX	0,004 02	0,006 38	0,006 29	0,006 05
	(0,002 0)	(0,112)	(0,037)	(0,035)
EDUC	0,000 58	0,008 96	0,004 65	0,005 49
	(0,000 3)	(0,327)	(0,032)	(0,025)
$\sigma_{v\alpha_1^*}$	0,000 30	0,017 35	0,021 15	0,015 80
	(0,002 2)	(3,243)	(0,272)	(0,179)
$\sigma^2_{\varepsilon_1}$	0,000 565	0,000 565	0,000 565	0,000 565
$\sigma^2_{\alpha_1}$	0,630 4	2,606 3	0,618 1	0,630 4

(*) Les écart-types sont entre parenthèses.

comme le taux de rendement de l'éducation, est positif. Cela signifie qu'une année d'études supérieures supplémentaire entraîne une hausse du salaire, toutes choses égales par ailleurs, de 0,9, 0,4 et 0,5 % pour HT, AM et BMS respectivement.

L'estimation d'un modèle avec biais de sélection indique que les instruments proposés par HT, à savoir ($Q_{M_1}\dot{X}_1$, $P_{M_1}\dot{X}_1$, $\dot{Z}_1$) fournissent des écart-types de l'ordre de 0,327, 0,311 et 0,010 pour EDUC, EXP et EXP^2 respectivement. Ces derniers passent à 0,032, 0,149 et 0,004 en utilisant les instruments (Q_{M_1}, $P_{M_1}\dot{X}_1$, $Q_{M_1}\dot{X}_1^0$, $\dot{Z}_1$) proposées par AM. Notons que HT utilisent chaque variable dans $\dot{X}_1$ comme deux instruments ($Q_{M_1}\dot{X}_1$ et $P_{M_1}\dot{X}_1$) alors que AM utilisent chacune de ces variables comme $(T+1)$ instruments ($Q_{M_1}\dot{X}_1$ et $\dot{X}_1^0$) où chaque colonne de $\dot{X}_1^0$ contient les valeurs de $\dot{X}_{1\,it}$ pour chaque période $t=1, \ldots, T$. On doit noter ici que cette différence de gains d'efficacité[7] enregistrée entre HT et AM est beaucoup plus importante que celle enregistré dans le cadre d'un modèle à un seul régime sans biais de sélection (CORNWELL et RUPERT [1988], BALTAGI et KHANTI-AKOUM [1990]). Cependant, on peut enregistrer la même différence, en comparant la première et la troisième colonne dans le tableau 3, entre les estimateurs de HT et BMS où les écarts-types passent à 0,025, 0,149 et

7. Il s'agit en fait d'une efficacité estimée.

0,004 pour les mêmes variables EDUC, EXP et EXP^2, alors que cette différence est beaucoup moins importante entre AM et BMS.

5 Conclusion

Dans cet article, nous avons présenté une procédure d'estimation d'un modèle à équations simultanées avec biais de sélection dans le cas de données de panel. L'approche de HT, permettant de traiter le cas des modèles et erreurs composées, est combinée avec celle de LMT, portant sur les équations simultanées avec auto-sélection, pour obtenir des estimateurs d'un modèle à variables explicatives constantes et variantes dans le temps. La matrice de variance-covariance asymptotique des paramètres est calculée en prenant en compte les modifications liées au cas des données de panel.

L'application au calcul des rendements de l'éducation permet de mesurer les performances relatives des matrices d'instruments de HT, AM et BMS. Pour l'estimation du premier régime, les méthodes de AM et BMS fournissent des estimateurs nettement plus efficaces que celle de HT. Par contre, la différence entre les instruments de AM et BMS est négligeable.

Estimation de $\Omega_1^{-1/2}$ et $\Omega_2^{-1/2}$

On a

$$\Omega_j^{-1/2} = Q_{M_j} + \theta_j P_{M_j}, \quad \theta_j = \frac{\sigma_{\varepsilon_j}}{(\sigma^2_{\varepsilon_j} + T\,\sigma^2_{\alpha_j})^{1/2}}, \qquad j = 1, 2$$

L'estimation de $\Omega_j^{-1/2}$ exige alors le calcul d'un estimateur convergent de θ_j, c'est-à-dire de $\sigma^2_{\alpha_j}$ et $\sigma^2_{\varepsilon_j}$. HT propose un estimateur convergent de $\sigma^2_{\varepsilon_j}$ basé sur une transformation des équations [6] et [7] par la matrice des projecteurs Q_{M_j}. Les effets individuels étant ainsi éliminés, il n'y a plus de corrélation entre ces derniers et le terme de perturbation v. Nous pouvons donc utiliser cet estimateur dans le cas présent.

Dans ce qui suit, on ne considérera que l'estimation de $\sigma^2_{\alpha_1}$, la méthode étant identique pour $\sigma^2_{\alpha_2}$. Soit $\hat{\beta}_1$ l'estimateur within-groups (intra-groupes) appliqué à l'équation (6). Pour obtenir un estimateur convergent de $\sigma^2_{\alpha_1}$, on définit, suivant HT, le vecteur de résidus $\hat{h}_1$:

(A.1) $$\hat{h}_1 = P_{M_1} Y_1 - P_{M_1} X_1 \hat{\beta}_1 \qquad \text{si} \quad i \in I_1$$

En remplaçant $\hat{\beta}_1$ par son expression, on a

(A.2) $$\hat{h}_1 = P_{M_1} Y_1 - P_{M_1} X_1 [X_1' Q_{M_1} X_1]^{-1} X_1' Q_{M_1} Y_1 \qquad \text{si} \quad i \in I_1$$

Après simplification, on obtient

(A.3) $$\hat{h}_1 = Z_1 \gamma_1 + \alpha_1 + [P_{M_1} - P_{M_1} X_1 [X_1' Q_{M_1} X_1]^{-1} X_1' Q_{M_1}] \varepsilon_1 \qquad \text{si} \quad i \in I_1$$

Considérons à présent l'estimation de γ_1 dans (A.3). Puisque α_1 est corrélé avec $\ddot{Z}_1$ et v, il faut calculer l'espérance de α_1 après avoir transformé les données par la matrice :[8]

$$P_{A_1} = A_1 (A_1' A_1)^{-1} A_1$$

où A_1 est la matrice des instruments (pour HT par exemple, elle est égale à $A_1 = [Q_{M_1}, \dot{X}_1, \dot{Z}_1]$).

On a

(A.4) $$E[\tilde{h}_1 \mid i \in I_1] = \tilde{Z}_1 \gamma_1 - \sigma^2_{v\tilde{\alpha}_1} \eta_1$$

8. Dans le cas d'un modèle à un seul régime, HT traitent les deux derniers termes de (A.3) comme une innovation inobservable de moyenne 0. Mais puisque α_1 est corrélé avec une partie des $\dot{Z}_1$, HT suggèrent une estimation de γ_1 par la méthode des doubles moindres carrés, avec la matrice d'instruments A_1.

où $\eta_1 = \frac{\phi}{\Phi}$, $\tilde{h}_1 = P_{A_1} \hat{h}_1$, $\tilde{Z}_1 = P_{A_1} Z_1$, $\tilde{\alpha}_1 = P_{A_1} \alpha_1$.

L'estimateur MCO de γ_1 dans (A.4) est calculé par la formule :

$$\text{(A.5)} \qquad \hat{\gamma}_1 = (\tilde{Z}_1' P_{A_1} R_1 P_{A_1} \tilde{Z}_1)^{-1} \tilde{Z}_1' P_{A_1} R_1 P_{A_1} \tilde{h}_1$$

où $R_1 = I_{TN_1} - \eta_1 (\eta_1' \eta_1)^{-1} \eta_1'$

Soit

$$\text{(A.6)} \quad D_1^2 = \frac{1}{N_1} (P_{M_1} Y_1 - P_{M_1} X_1 \hat{\beta}_1 - Z_1 \hat{\gamma}_1)' (P_{M_1} Y_1 - P_{M_1} X_1 \hat{\beta}_1 - Z_1 \hat{\gamma}_1)$$

Il vient :

$$\lim_{N_1 \to \infty} D_1^2 = \lim_{N_1 \to \infty} \frac{1}{N_1} (\alpha_1 + P_{M_1} \varepsilon_1)' (\alpha_1 + P_{M_1} \varepsilon_1) = \sigma_{\alpha_1}^2 + \frac{1}{T} \sigma_{\varepsilon_1}^2$$

Finalement,

$$\text{(A.7)} \qquad \hat{\sigma}_{\alpha_1}^2 = D_1^2 - \frac{1}{T} \hat{\sigma}_{\varepsilon_1}^2$$

est un estimateur convergent de $\sigma_{\alpha_1}^2$.

• Références bibliographiques

AMEMIYA, T. (1973). – « Regression Analysis when the Dependent Variable is Truncated », *Econometrica*, 41, pp. 997-1016.

AMEMIYA, T. (1974). – « Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal », *Econometrica*, 42, pp. 999-1012.

AMEMIYA, T. (1978). – « The Estimation of a Simultaneous Equation Generalized Probit Model », *Econometrica*, 46, pp. 1193-1205.

AMEMIYA, T. et MCCURDY, T. (1986). – « Instrumental Variable Estimation of an Error-Component Model », *Econometrica*, 54, pp. 869-880.

BALESTRA, P. et NERLOVE, M. (1966). – « Pooling Cross-Section and Time-Series Data in Estimation of a Dynamic Model: The Demand for Natural Gas », *Econometrica*, 34, pp. 585-616.

BALTAGI, B. H. et KHANTI-AKOUM, S. (1990). – « On Efficient Estimation with Panel data: An Empirical Comparison of Instrumental Variables Estimators », *Journal of Applied Econometrics*, 5, pp. 401-406.

BOUMAHDI, R. (1991 *a*). – « Les rendements de l'éducation : analyse et problèmes économétriques », Thèse de doctorat en Sciences Économiques, Université de Toulouse I.

BOUMAHDI, R. (1991 *b*). – « Consistent Estimation of a Two Equation-Error Components Model », Cahier *GREMAQ* 9011, présenté à l'Econometric Society European Meeting, Cambridge.

Boumahdi, R., Florens, J. P. et Plassard, J. M. (1990). – « Estimation des rendements de l'éducation : Un modèle à variable dépendante tronquée, *Document de Travail présenté aux 7e Journées de Microéconomie Appliquée*, Montréal.

Boumahdi, R. et Plassard, J. M. (1992). – « Note à propos du caractère endogène de la variable éducation dans la fonction de gain », *Revue Économique*, 43, pp. 145-156.

Breusch, T. S., Mizon, G. E. et Schmidt, P. (1989). – « Efficient Estimation Using Panel Data », *Econometrica*, 57, pp. 695-700.

Chamberlain, G. et Griliches, Z. (1975). – « Unobservables with a Variance-Components Structure: Ability, Schooling and the Economics Sucess of Brothers », *International Econometric Review*, 16, pp. 422-449.

Cornwell, C. et Ruppert, P. (1988). – « Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables Estimators », *Journal of Applied Econometrics*, 3, pp. 149-155.

Griliches, Z. (1977). – « Estimating the Returns to Schooling: Some Econometric Problems », *Econometrica*, 45, pp. 1-22.

Hausman, J. A. (1978). – « Specification Tests in Econometrics », *Econometrica*, 46, pp. 1251-1271.

Hausman, J. A. et Taylor, W. E. (1981). – « Panel Data and Unobservable Individual Effects », *Econometrica*, 49, pp. 1377-1398.

Heckman, J. J. (1976). – « The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dpendent Variables and a Simple Estimator for Such Models », *Annals of Economic and Social Measurement*, 5, pp. 475-492.

Heckman, J. J. (1979). – « Sample Selection Bias as a Specification Error », *Econometrica*, 47, pp. 153-161.

INSEE (1982). – « Recensement Général de la Population, France Métropolitaine. Structure de la Population Totale », Série D, RP82.

INSEE (1990). – *Annuaire Statistique de la France*, Résultats de 1989.

Lee, L. F. et Trost, R. P. (1978). – « Estimation of Some Limited Dependent Variable Models with Applications to Housing Demand », *Journal of Econometrics*, 8, pp. 357-382.

Lee, L. F., Maddala, G. S. et Trost, R. P. (1980). – « Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity », *Econometrica*, 48, pp. 491-503.

Maddala, G. S. (1971). – « The Use of Variance Components Models in Pooling Cross Section and Time Series Data », *Econometrica*, 39, pp. 341-358.

Nelson, F. D. et Olson, L. (1978). – « Specification and Estimation of a Simultaneous Equations Model with Limited Dependent Variables », *International Economic Review*, 19, pp. 695-709.

Nerlove, M. (1971). – « Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections », *Econometrica*, 39, pp. 359-382.

Tobin, J. (1958). – « Estimation of Relationship for Limited Dependent Variables », *Econometrica*, 26, pp. 24-36.

Willis, R. J. et Rosen, S. (1979). – « Education and Self-Selection ». *Journal of Political Economy*, 87, s7-s36.