

ENDOGENÉITÉ DANS UN SYSTÈME D'ÉQUATIONS NORMAL BIVARIÉ AVEC VARIABLES QUALITATIVES

Stéfan LOLLIVIER

ENSAE

L'auteur remercie Christian GOURIEROUX pour ses remarques sur une version antérieure du texte. Les erreurs qui pourraient subsister relèvent de la responsabilité de l'auteur.

Introduction

La prise en compte de l'endogénéité d'une variable explicative conduit à expliciter puis à estimer un système de deux équations simultanées se présentant sous forme structurelle. La plupart du temps, il en résulte que l'estimation de l'une des équations sans tenir compte de l'autre conduit en général 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 n'est plus nulle. Cependant, le traitement de l'endogénéité des variables explicatives est maintenant bien établi lorsque la variable expliquée comme la variable explicative sont continues (doubles moindres carrés,...). Parmi les approches disponibles, la méthode des régressions augmentées permet de réaliser à moindres frais des tests de spécification mais aussi d'effectuer des estimations permettant de corriger de tels biais (voir par exemple [6]). On dispose ainsi de procédures efficaces 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 faisant appel à des variables qualitatives. Celles-ci peuvent apparaître en tant que variables explicatives ou comme variables expliquées. La résolution du modèle à équations simultanées qui en résulte est alors beaucoup plus délicate. On dispose bien sûr de la méthode du maximum de vraisemblance, qui fournit des estimateurs asymptotiquement convergents et permet de tester l'endogénéité des variables. Mais cette technique suppose en général dans la pratique d'avoir recours à une programmation spécifique. Cette difficulté peut être contournée dans un certain nombre de cas. Heckman [2] a ainsi présenté une formalisation générale permettant de traiter le problème d'endogénéité lorsque la variable expliquée est continue et la variable explicative dichotomique. Il est en effet possible d'avoir recours à une procédure en deux étapes, la première consistant en l'estimation d'un modèle probit portant sur la variable qualitative suspecte d'endogénéité, la seconde en une régression linéaire augmentée d'un terme issu de la première étape. L'estimateur des moindres carrés ordinaires portant sur cette régression est alors convergent et asymptotiquement normal ([1]). La complication pratique consiste seulement à réaliser de façon rigoureuse le test permettant de conclure ou non à l'endogénéité, la

matrice de variance-covariance des estimateurs n'étant pas celle qui ressort de l'estimation des moindres carrés ordinaires, en raison de l'hétéroscédasticité du terme d'erreur présent dans cette régression augmentée.

La complexité s'accroît encore d'un degré lorsque la variable expliquée est elle aussi qualitative. Même si les développements classiques restent vrais sur les variables latentes sous-jacentes aux variables dichotomiques, les problèmes pratiques liés à l'estimation se posent dès lors que l'on n'a plus affaire à des densités mais à des probabilités. S'il est encore possible d'avoir recours à des méthodes faisant intervenir la maximisation d'une vraisemblance conditionnelle lorsque la variable explicative est quantitative, le recours à une procédure de maximisation de la vraisemblance est pratiquement incontournable lorsque les deux variables sont qualitatives.

1. Rappels concernant le cas où les deux variables sont observées

1.1 Le modèle

On considère le modèle bivarié suivant, dans lequel les variables y_{1i} et y_{2i} sont observées:

$$\begin{cases} y_{1i} &= X_{1i} \mathbf{b}_1 + \mathbf{g} y_{2i} + u_{1i} \\ y_{2i} &= X_{2i} \mathbf{b}_2 + u_{2i} \end{cases},$$

Les termes d'erreur sont indépendants et identiquement distribués entre les individus. Si on les suppose seulement d'espérance nulle, le modèle n'est identifiable (notamment \mathbf{b}_1 et \mathbf{b}_2) que si X_{1i} diffère de X_{2i} . Si on postule des conditions supplémentaires sur les moments d'ordre supérieur à 1, ce qui est généralement le cas lorsque l'on suppose que les perturbations ont une variance donnée, et a fortiori si l'on choisit une loi sur les perturbations, le modèle devient identifiable même si X_{1i} et X_{2i} sont égaux. En pratique, on s'efforcera cependant dans la mesure du possible de faire en sorte que les deux familles de variables explicatives soient distinctes. En effet, il est préférable que le modèle soit identifiable à l'ordre un, ce qui correspond à une résolution du système en l'absence de termes d'erreurs, et donc à la représentation sous-jacente du phénomène.

On suppose en outre que l'espérance de u_{1i} conditionnelle à X_{1i}, X_{2i} et u_{2i} est linéaire en u_{2i} , c'est à dire que l'on peut écrire u_{1i} sous la forme :

$$u_{1i} = \mathbf{r} \frac{\mathbf{s}_1}{\mathbf{s}_2} u_{2i} + v_{1i},$$

v_{1i} étant une variable aléatoire indépendante de u_{2i} telle que $E(v_{1i}) = 0$. On en déduit que :

$$V(v_{1i}) = \mathbf{s}_1^2 (1 - \mathbf{r}^2),$$

Avec un tel modèle,

$$E(y_{1i} / X_{1i}, X_{2i}, y_{2i}) = X_{1i} \mathbf{b}_1 + \mathbf{g} y_{2i} + E(u_{1i} / X_{1i}, X_{2i}, y_{2i}),$$

soit, compte-tenu des résultats précédents :

$$E(y_{1i} / X_{1i}, X_{2i}, y_{2i}) = X_{1i} \mathbf{b}_1 + \mathbf{g} y_{2i} + \mathbf{r} \frac{\mathbf{s}_1}{\mathbf{s}_2} (y_{2i} - X_{2i} \mathbf{b}_2)$$

Cette expression fait immédiatement apparaître que l'estimation isolée de la première équation par les moindres carrés ordinaires conduit à des estimateurs biaisés et non convergents dès lors que $r \neq 0$, en raison de l'endogénéité de la variable y_2 .

1.2 L'estimation

On peut néanmoins réécrire la première équation du modèle sous la forme :

$$y_{1i} = X_{1i} \mathbf{b}_1 + \mathbf{g} y_{2i} + \mathbf{r} \frac{\mathbf{s}_1}{\mathbf{s}_2} (y_{2i} - X_{2i} \mathbf{b}_2) + v_{1i}$$

avec $E(v_{1i} / X_{1i}, X_{2i}, y_{2i}) = 0$, $V(v_{1i} / X_{1i}, X_{2i}, y_{2i}) = \mathbf{s}_1^2 (1 - r^2)$, les termes d'erreur v_{1i} étant toujours indépendants entre eux.

L'estimation de la seconde équation par les moindres carrés ordinaires permet d'obtenir pour chaque observation un résidu $\hat{u}_{2i} = y_{2i} - X_{2i} \hat{\mathbf{b}}_2$. On en déduit :

$$y_{1i} = X_{1i} \mathbf{b}_1 + \mathbf{g} y_{2i} + \mathbf{r} \frac{\mathbf{s}_1}{\mathbf{s}_2} \hat{u}_{2i} + v_{1i} + \mathbf{r} \frac{\mathbf{s}_1}{\mathbf{s}_2} (u_{2i} - \hat{u}_{2i}) \quad (1)$$

avec $E\left(v_{1i} + \mathbf{r} \frac{\mathbf{s}_1}{\mathbf{s}_2} (u_{2i} - \hat{u}_{2i})\right) = 0$ et toujours $V(v_{1i}) = \mathbf{s}_1^2 (1 - r^2)$.

En revanche, le fait d'avoir remplacé u_{2i} par l'estimateur \hat{u}_{2i} conduit à l'apparition d'un nouveau terme d'erreur dans l'équation :

$$\mathbf{r} \frac{\mathbf{s}_1}{\mathbf{s}_2} (u_{2i} - \hat{u}_{2i})$$

Ce dernier est bien d'espérance nulle, mais si l'on considère l'ensemble de l'échantillon,

$$V(u_2 - \hat{u}_2) = V(X_2 (\hat{\mathbf{b}}_2 - \mathbf{b}_2)) = \mathbf{s}_2^2 X_2 (X_2' X_2)^{-1} X_2'$$

Il en résulte une hétéroscédasticité au niveau de chaque observation, mais surtout une corrélation entre les termes d'erreur relatifs aux différentes observations : la matrice de variance-covariance des $u_{2i} - \hat{u}_{2i}$ n'est pas diagonale, et par conséquent le modèle ne respecte plus l'indépendance des termes d'erreur entre les observations.

Estimer l'équation augmentée (1) par les moindres carrés ordinaires fournit des estimateurs convergents et asymptotiquement normaux, mais il faut calculer leur matrice de variance-covariance en tenant compte de ces caractéristiques particulières. En l'espèce, on sait que l'estimateur des paramètres ainsi obtenu est celui des doubles moindres carrés, la seconde équation du modèle simultané pouvant se réinterpréter comme une régression instrumentale [6]. La matrice de variance-covariance est donc égale à :

$$V\left(\begin{matrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{g}} \end{matrix}\right) = (\bar{X}' P_2 \bar{X})^{-1}$$

avec

$$\bar{X} = (X_1; y_{2i}), \text{ et } P_2 = X_2 (X_2' X_2)^{-1} X_2'$$

2. Variable expliquée qualitative, variable explicative continue

2.1 La maximisation de la vraisemblance

On n'observe cette fois qu'une représentation discrète y_{1i} de la variable expliquée latente y_{1i}^* , en deux ou plusieurs tranches ; la variable explicative y_{2i} est pour sa part observée et continue. On aboutit au modèle latent bivarié suivant :

$$\begin{cases} y_{1i}^* &= X_{1i} \mathbf{b}_1 + \mathbf{g} y_{2i} + u_{1i} \\ y_{2i} &= X_{2i} \mathbf{b}_2 + u_{2i} \end{cases}$$

Afin de pouvoir calculer des estimateurs du maximum de vraisemblance, il faut postuler une loi sur les perturbations. On effectue une hypothèse de binormalité, de sorte que :

$$\begin{pmatrix} u_{1i} \\ u_{2i} \end{pmatrix} \rightarrow N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{s}_1^2 & r \mathbf{s}_1 \mathbf{s}_2 \\ r \mathbf{s}_1 \mathbf{s}_2 & \mathbf{s}_2^2 \end{pmatrix} \right]$$

Même si cela n'est pas indispensable en matière d'identification, on évitera à nouveau que X_{1i} et X_{2i} soient égaux. Sinon, les restrictions habituelles aux modèles qualitatifs, notamment celles portant sur \mathbf{s}_1 , permettent d'assurer l'identification du modèle, selon qu'il s'agit d'un modèle dichotomique ou polytomique ordonné, les seuils étant ou non connus. Ainsi, dans le cas d'un modèle dichotomique ou d'un modèle polytomique ordonné à seuils inconnus, le paramètre \mathbf{s}_1 doit être contraint à 1. En revanche, si les seuils sont connus (cas d'une variable en tranches), une telle contrainte n'est pas nécessaire.

L'hypothèse de binormalité permet d'écrire u_{1i} sous la forme :

$$u_{1i} = r \frac{\mathbf{s}_1}{\mathbf{s}_2} u_{2i} + v_{1i}$$

avec $E(v_{1i} / X_{1i}, X_{2i}, y_{2i}) = 0$, et $V(v_{1i} / X_{1i}, X_{2i}, y_{2i}) = \mathbf{s}_1^2 (1 - r^2)$, v_{1i} étant une variable aléatoire normale indépendante de u_{2i} . Le modèle latent peut alors s'écrire :

$$y_{1i}^* = X_{1i} \mathbf{b}_1 + \mathbf{g} y_{2i} + r \frac{\mathbf{s}_1}{\mathbf{s}_2} (y_{2i} - X_{2i} \mathbf{b}_2) + v_{1i}$$

La variable y_2 étant continue, la contribution d'un individu i à la vraisemblance dans le cas où y_{1i} est dichotomique, s'écrit :

$$L_i = Z_{1i} \left[\frac{1}{\mathbf{s}_1 \sqrt{1 - r^2}} \left(X_{1i} \mathbf{b}_1 + \mathbf{g} y_{2i} + \frac{r}{\mathbf{s}_2} \mathbf{s}_1 (y_{2i} - X_{2i} \mathbf{b}_2) \right) \right] \frac{1}{\mathbf{s}_2} \mathbf{j} \left(\frac{y_{2i} - X_{2i} \mathbf{b}_2}{\mathbf{s}_2} \right)$$

où $Z_{1i}(x) = \Phi(x)$ si $Y_{1i} = 1$ et $Z_{1i}(x) = 1 - \Phi(x)$, si $Y_{1i} = 0$.

Dans les logiciels usuels, il n'existe pas en général de procédure intégrée pour déterminer les estimateurs du maximum de vraisemblance correspondant à ce type de vraisemblance. En revanche, la plupart d'entre eux met à la disposition des utilisateurs un algorithme général de maximisation de fonctions. Il reste à l'utilisateur à programmer «à la main» la vraisemblance, et faire en sorte au travers de différents réglages que la maximisation se déroule convenablement, ce qui est beaucoup moins confortable que d'utiliser une procédure presse-bouton. Ceci étant, en l'espèce, la fonction de

répartition de la loi normale est disponible dans tous les logiciels statistiques, et la programmation de cette vraisemblance ne soulève guère de difficultés.

2.2 Une procédure en deux étapes

Il est cependant possible, au prix d'une moindre efficacité des estimateurs, de mettre en œuvre une procédure en deux étapes, qui utilise les fonctionnalités standard des logiciels, et ne nécessite pas de recourir à une programmation trop ad hoc. On a vu que sous l'hypothèse de binormalité de la perturbation, la première équation pouvait se réécrire sous la forme :

$$y_{1i}^* = X_{1i} \mathbf{b}_1 + \mathbf{g} y_{2i} + \mathbf{r} \frac{\mathbf{s}_1}{\mathbf{s}_2} u_{2i} + v_{1i} \quad (2)$$

dans laquelle v_{1i} est une perturbation i.i.d., normale et homoscedastique. Si u_{2i} était observé, le modèle pourrait s'estimer la méthode du maximum de vraisemblance comme un modèle probit simple. Ici, on ne dispose pas de la valeur de u_{2i} , mais l'estimation de la seconde équation par les moindres carrés ordinaires en fournit un estimateur sans biais, que l'on peut substituer à la valeur inconnue. On sait par ailleurs que cette substitution permet d'obtenir pour N infini des estimateurs convergents et asymptotiquement normaux des paramètres présents dans l'équation [5].

L'estimation de l'équation (2), dans laquelle on a remplacé u_{2i} par \hat{u}_{2i} , au moyen d'un modèle probit simple par un logiciel standard fournit donc des estimateurs sans biais et asymptotiquement normaux des paramètres. Leur matrice de variance-covariance asymptotique diffère cependant de ce que fournit le logiciel standard, en raison du remplacement de la variable inconnue par l'imputation. Un moyen commode pour déterminer un estimateur des écart-types des paramètres est d'utiliser une technique de bootstrap. L'efficacité des estimateurs en deux étapes sera cependant moindre que celle obtenue par la méthode du maximum de vraisemblance décrite en 2.1.

3. La variable explicative est qualitative, la variable expliquée observée

3.1 Le cas où la variable y_2 est dichotomique.

3.1.1 l'estimateur du maximum de vraisemblance

On observe les variables explicatives X_{1i} et X_{2i} , la variable y_{1i} ainsi que la variable muette d_i qui correspond à la positivité de la variable latente y_{2i}^* . Le système s'écrit par conséquent :

$$\begin{cases} y_{1i} &= X_{1i} \mathbf{b}_1 + \mathbf{a} d_i + u_{1i} \\ y_{2i}^* &= X_{2i} \mathbf{b}_2 + u_{2i} \end{cases}$$

avec une hypothèse sur les lois des perturbations (binormalité) nécessaire au calcul de la vraisemblance :

$$\begin{pmatrix} u_{1i} \\ u_{2i} \end{pmatrix} \rightarrow N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{s}_1^2 & \mathbf{r} \mathbf{s}_1 \mathbf{s}_2 \\ \mathbf{r} \mathbf{s}_1 \mathbf{s}_2 & \mathbf{s}_2^2 \end{pmatrix} \right]$$

Afin de rendre ce modèle identifiable, on introduit la contrainte $\mathbf{s}_2 = 1$. La contribution à la vraisemblance d'un individu i a alors pour expression :

$$L_i = Z_{2i} \left[\frac{1}{\sqrt{1-r^2}} \left(X_{2i} \mathbf{b}_2 + \frac{\mathbf{r}}{\mathbf{s}_1} (y_{1i} - X_{1i} \mathbf{b}_1 - \mathbf{a} d_i) \right) \right] \frac{1}{\mathbf{s}_1} \mathbf{j} \left(\frac{y_{1i} - X_{1i} \mathbf{b}_1 - \mathbf{a} d_i}{\mathbf{s}_1} \right)$$

où $Z_{2i}(x) = \Phi(x)$ si $Y_{2i} = 1$ et $Z_{2i}(x) = 1 - \Phi(x)$, si $Y_{2i} = 0$.

A nouveau, la détermination des estimateurs du maximum de vraisemblance ne soulève pas de réelles difficultés numériques, mais nécessite le recours à une programmation spécifique dans la mise en œuvre de l'algorithme itératif général de maximisation d'une fonction.

3.1.2 L'estimateur en deux étapes

Comme dans la deuxième section, on peut mettre en œuvre une procédure en deux étapes plus économe en coûts de programmation. Celle-ci présente en outre l'avantage de pouvoir alléger l'hypothèse de binormalité requise par l'estimation du maximum de vraisemblance, en particulier concernant la perturbation relative à la première équation linéaire. On suppose seulement la normalité de la perturbation u_{2i} et comme dans la section 1, le fait que conditionnellement aux variables explicatives, la régression de u_{1i} sur u_{2i} est linéaire, ce qui peut s'écrire :

$$u_{1i} = \mathbf{r} \mathbf{s}_1 u_{2i} + v_{1i},$$

avec $E(v_{1i} / X_{1i}, X_{2i}, d_i) = 0$, et $V(v_{1i} / X_{1i}, X_{2i}, d_i) = \mathbf{s}_1^2 (1 - r^2)$. Cette hypothèse est vérifiée en cas de binormalité, comme conséquence des propriétés des lois normales, mais alors v_{1i} est lui-aussi normal. Pour le calcul de l'estimateur en deux étapes, cette hypothèse de normalité de v_{1i} n'est pas nécessaire, ce qui permet d'alléger les contraintes paramétriques nécessaires pour l'identification du modèle.

Sous ces hypothèses, on peut montrer que :

$$E(y_{1i} / X_{1i}, X_{2i}, d_i = 1) = X_{1i} \mathbf{b}_1 + \mathbf{a} + \mathbf{r} \mathbf{s}_1 \mathbf{I}_i, \text{ avec } \mathbf{I}_i = \frac{\mathbf{j}(X_{2i} \mathbf{b}_2)}{\Phi(X_{2i} \mathbf{b}_2)}$$

et

$$E(y_{1i} / X_{1i}, X_{2i}, d_i = 0) = X_{1i} \mathbf{b}_1 + \mathbf{a} + \mathbf{r} \mathbf{s}_1 \tilde{\mathbf{I}}_i, \text{ avec } \tilde{\mathbf{I}}_i = \frac{-\mathbf{j}(X_{2i} \mathbf{b}_2)}{1 - \Phi(X_{2i} \mathbf{b}_2)},$$

soit pour $d_i = 1$,

$$y_{1i} = X_{1i} \mathbf{b}_1 + \mathbf{a} + \mathbf{r} \mathbf{s}_1 \mathbf{I}_i + v_{1i},$$

v_{1i} étant une variable aléatoire d'espérance nulle et de variance $V(v_{1i}) = \mathbf{s}_1^2 - \mathbf{s}_1^2 r^2 A_i$, avec $A_i = (X_{2i} \mathbf{b}_2) \mathbf{I}_i + \mathbf{I}_i^2$ et pour $d_i = 0$,

$$y_{1i} = X_{1i} \mathbf{b}_1 + \mathbf{a} + \mathbf{r} \mathbf{s}_1 \tilde{\mathbf{I}}_i + v_{1i},$$

v_{1i} étant une variable aléatoire d'espérance nulle et de variance $V(v_{1i}) = \mathbf{s}_1^2 - \mathbf{s}_1^2 r^2 B_i$, avec $B_i = -(X_{2i} \mathbf{b}_2) \tilde{\mathbf{I}}_i + \tilde{\mathbf{I}}_i^2$.

Au total, si l'on note $\mathbf{m}_i = \mathbf{I}_i d_i + \tilde{\mathbf{I}}_i (1 - d_i)$ et $C_i = A_i d_i + B_i (1 - d_i)$, le modèle s'écrit :

$$y_{li} = X_{li} \mathbf{b}_1 + \mathbf{a} d_i + \mathbf{r} \mathbf{s}_1 \mathbf{m}_i + v_{li}$$

\bar{v}_{li} étant une variable aléatoire d'espérance nulle conditionnellement à d_i et de variance $\mathbf{s}_1^2 - \mathbf{s}_1^2 \mathbf{r}^2 C_i$. Les termes d'erreur sont en outre indépendants entre les observations ([3]).

Par rapport au cas où la variable explicative est continue, le modèle ainsi réécrit permet à nouveau de prendre en compte l'endogénéité, mais au prix d'un certain nombre de réserves. D'une part, le terme d'erreur \bar{v}_{li} n'est pas normal. D'autre part, sa variance n'est pas la même pour tous les individus, le terme C_i étant différent d'une observation à l'autre. Contrairement aux sections 1 et 2, l'hétéroscédasticité est intrinsèque à ce modèle et ne provient que du remplacement d'une variable inobservée par une variable imputée.

Il est possible d'estimer la seconde équation au moyen d'un modèle probit ; on obtient alors un estimateur convergent de \mathbf{b}_2 , qui peut servir à calculer les estimateurs $\hat{\mathbf{I}}_i$ et $\hat{\tilde{\mathbf{I}}}_i$ de \mathbf{I}_i et $\tilde{\mathbf{I}}_i$ suivants :

$$\hat{\mathbf{I}}_i = \frac{\mathbf{j}(X_{2i} \hat{\mathbf{b}}_2)}{\Phi(X_{2i} \hat{\mathbf{b}}_2)} \text{ et } \hat{\tilde{\mathbf{I}}}_i = \frac{-\mathbf{j}(X_{2i} \hat{\mathbf{b}}_2)}{1 - \Phi(X_{2i} \hat{\mathbf{b}}_2)}$$

D'où la nouvelle équation :

$$y_{li} = X_{li} \mathbf{b}_1 + \mathbf{a} d_i + \mathbf{r} \mathbf{s}_1 \hat{\mathbf{m}}_i + w_{li} \quad (3)$$

avec $\hat{\mathbf{m}}_i = \hat{\mathbf{I}}_i d_i + \hat{\tilde{\mathbf{I}}}_i (1 - d_i)$, et $w_{li} = v_{li} + \mathbf{r} \mathbf{s}_1 (\mathbf{m}_i - \hat{\mathbf{m}}_i)$.

On peut noter que w_{li} est un terme d'erreur d'espérance nulle.

L'approximation au premier ordre :

$$\mathbf{m}_i - \hat{\mathbf{m}}_i \approx C_i X_{2i} (\hat{\mathbf{b}}_2 - \mathbf{b}_2),$$

permet de déterminer la matrice de variance-covariance du terme d'erreur additionnel :

$$V(\mathbf{m} - \hat{\mathbf{m}}) \approx \mathbf{C} X_2 V(\hat{\mathbf{b}}_2) \mathbf{C} X_2'$$

où $\mathbf{C} X_2$ est la matrice dont les lignes sont les $C_i X_{2i}$.

A nouveau, le fait de remplacer dans la régression un terme inobservé par un terme imputé introduit une perturbation entraînant une corrélation entre les termes d'erreur des différentes observations. Ceci ne gêne en rien le calcul de l'estimateur des moindres carrés ordinaires, qui est convergent et asymptotiquement normal. Ceci complexifie en revanche le calcul de sa matrice de variance-covariance asymptotique. Lee, Maddala et Trost [3] montrent au prix de longs calculs, que celle-ci s'écrit :

$$V(\hat{\mathbf{b}}) = \mathbf{s}_1^2 (\mathbf{G}' \mathbf{G})^{-1} - \mathbf{s}_1^2 (1 - \mathbf{r}^2) (\mathbf{G}' \mathbf{G})^{-1} \mathbf{G}' [\mathbf{C} - \mathbf{C} X_2 V(\hat{\mathbf{b}}_2) \mathbf{C} X_2'] \mathbf{G} (\mathbf{G}' \mathbf{G})^{-1}$$

où $\mathbf{G} = (X_1; d; \hat{\mathbf{m}})$, et $V(\hat{\mathbf{b}}_2)$ est la variance de l'estimateur des paramètres estimés par le modèle probit dans la seconde équation. Cette expression montre en outre l'erreur qui serait commise dans le calcul de cette variance-covariance si on omettait de tenir compte du fait qu'une variable est imputée dans l'estimation.

Cette procédure en deux étapes (probit puis mco) permet d'obtenir des estimateurs convergents des paramètres d'intérêt et de leur matrice de variance-covariance asymptotique. L'estimateur des mco étant asymptotiquement normal, il est alors possible d'effectuer des tests asymptotiques, notamment le test d'endogénéité. Un autre moyen plus économe pour calculer cette matrice de variance-covariance est de recourir à des techniques de bootstrap.

Un cas analogue est celui pour lequel $\mathbf{a} = 0$ et où y_{li} n'est observé que si $d_i = 1$. Il correspond à la situation du modèle dit «Tobit généralisé». Sur le sous-échantillon où $d_i = 1$, l'équation (3) est toujours valide, puisque par construction $E(w_{li} / d_i) = 0$. La procédure en deux étapes décrite précédemment est alors rigoureusement celle définie par Heckman.

3.1.3 Un test de spécification

Le fait de disposer de deux estimateurs des paramètres de la première équation permet de mettre en œuvre un test de spécification «à la Haussman» portant sur la binormalité des termes d'erreurs dans l'estimation par le maximum de vraisemblance. Rappelons que ce test consiste à comparer deux estimateurs dont l'un, noté $\hat{\mathbf{d}}_1$ est convergent et asymptotiquement efficace sous l'hypothèse testée H_0 , mais non convergent sous H_1 , et l'autre, noté $\hat{\mathbf{d}}_2$, est convergent sous H_0 et H_1 . Le test consiste à rejeter H_0 si la statistique :

$$Q_H = (\hat{\mathbf{d}}_2 - \hat{\mathbf{d}}_1)' [\hat{V}(\hat{\mathbf{d}}_2) - \hat{V}(\hat{\mathbf{d}}_1)]^{-1} (\hat{\mathbf{d}}_2 - \hat{\mathbf{d}}_1)$$

est supérieure ou égale au fractile de la loi du χ^2 à $\dim(\mathbf{d})$ degrés de liberté, et à l'accepter sinon.

Dans le cas qui nous intéresse $\mathbf{d} = (\mathbf{b}_1; \mathbf{a}')$. L'estimateur $\hat{\mathbf{d}}_1$ est fourni par l'estimation du modèle par la méthode du maximum de vraisemblance, qui suppose la binormalité des termes d'erreur. L'estimateur $\hat{\mathbf{d}}_2$ correspond pour sa part à l'estimateur en deux étapes de ces paramètres. Sous l'hypothèse H_0 de binormalité, l'estimateur du maximum de vraisemblance est convergent et efficace, et l'estimateur en deux étapes est convergent. Sous H_1 , hypothèse moins restrictive de normalité de u_2 et de régression linéaire de u_1 sur u_2 , l'estimateur en deux étapes est convergent, contrairement à l'estimateur du maximum de vraisemblance.

3.2 Le cas où la variable y_2 est polytomique

Les développements précédents peuvent s'étendre à la situation où la variable y_2 est polytomique (les seuils étant connus ou non). Si l'on observe K états, le modèle bivarié peut s'écrire :

$$\begin{cases} y_{li} &= X_{li} \mathbf{b}_1 + \sum_{k=1}^K \mathbf{a}_k d_{ki} + u_{li} \\ y_{2i}^* &= X_{2i} \mathbf{b}_2 + u_{2i} \end{cases}$$

où $d_k = 1$ si l'on observe l'état k . On contraint alors l'un des \mathbf{a}_k à la nullité pour assurer l'identifiabilité. Sous les mêmes hypothèses portant sur les termes d'erreur :

$$E(y_{li} / X_{li}, X_{2i}, y_{2i}^*) = X_{li} \mathbf{b}_1 + \sum_{k=1}^K \mathbf{a}_k d_{ki} + \mathbf{r} \frac{\mathbf{s}_1}{\mathbf{s}_2} (y_{2i}^* - X_{2i} \mathbf{b}_2),$$

on en déduit :

$$E\left(y_{1i} / X_{1i}, X_{2i}; s_k \leq y_{2i}^* < s_{k+1}\right) = X_{1i} \mathbf{b}_1 + \mathbf{a}_k + \mathbf{r} \mathbf{s}_1 E\left(\frac{y_{2i}^* - X_{2i} \mathbf{b}_2}{\mathbf{s}_2} \middle/ \frac{s_k - X_{2i} \mathbf{b}_2}{\mathbf{s}_2} \leq \frac{y_{2i}^* - X_{2i} \mathbf{b}_2}{\mathbf{s}_2} < \frac{s_{k+1} - X_{2i} \mathbf{b}_2}{\mathbf{s}_2}\right)$$

$$= X_{1i} \mathbf{b}_1 + \mathbf{a}_k + \mathbf{r} \mathbf{s}_1 \frac{j\left(\frac{s_k - X_{2i} \mathbf{b}_2}{\mathbf{s}_2}\right) - j\left(\frac{s_{k+1} - X_{2i} \mathbf{b}_2}{\mathbf{s}_2}\right)}{\Phi\left(\frac{s_{k+1} - X_{2i} \mathbf{b}_2}{\mathbf{s}_2}\right) - \Phi\left(\frac{s_k - X_{2i} \mathbf{b}_2}{\mathbf{s}_2}\right)}$$

On peut alors écrire :

$$y_{1i} = X_{1i} \mathbf{b}_1 + \sum_{k=1}^K \mathbf{a}_k d_{ik} + \mathbf{r} \mathbf{s}_1 \sum_{k=1}^K \frac{j\left(\frac{s_k - X_{2i} \mathbf{b}_2}{\mathbf{s}_2}\right) - j\left(\frac{s_{k+1} - X_{2i} \mathbf{b}_2}{\mathbf{s}_2}\right)}{\Phi\left(\frac{s_{k+1} - X_{2i} \mathbf{b}_2}{\mathbf{s}_2}\right) - \Phi\left(\frac{s_k - X_{2i} \mathbf{b}_2}{\mathbf{s}_2}\right)} d_{ik} + v_{1i} \quad (3bis)$$

avec un terme d'erreur v_{1i} à nouveau centré et hétéroscédastique. Pour estimer ce modèle, et conduire des tests asymptotiques, on peut à nouveau utiliser la procédure en deux étapes décrite précédemment :

- un modèle probit polytomique permet d'estimer \mathbf{b}_2 ,
- puis l'estimation par les moindres carrés ordinaires de l'équation (3bis), dans laquelle on aura substitué $\hat{\mathbf{b}}_2$ à \mathbf{b}_2 permet d'obtenir $\hat{\mathbf{b}}_1$,
- et enfin la détermination d'un estimateur de la matrice de variance-covariance asymptotique de $\hat{\mathbf{b}}_1$ par une technique de bootstrap.

4. Modèle qualitatif dichotomique bivarié

4.1 Les limites d'un estimateur en deux étapes

Le modèle qualitatif dichotomique bivarié représente la configuration la plus complexe dans laquelle seuls les signes y_1 et $d_i = y_2$ des deux variables y_1^* et y_2^* sont observées :

$$\begin{cases} y_{1i}^* &= X_{1i} \mathbf{b}_1 + \mathbf{a} d_i + u_{1i}, \\ y_{2i}^* &= X_{2i} \mathbf{b}_2 + u_{2i}. \end{cases}$$

L'équation (3) demeure valide, et peut se réécrire :

$$y_{1i}^* = X_{1i} \mathbf{b}_1 + \mathbf{a} d_i + \mathbf{r} \mathbf{s}_1 \hat{\mathbf{m}}_i + w_{1i} \quad (4),$$

avec

$$w_{1i} = v_{1i} + \mathbf{r} \mathbf{s}_1 (\mathbf{m}_i - \hat{\mathbf{m}}_i)$$

représentant un terme d'erreur d'espérance nulle.

Mais contrairement au cas où y_1^* était observé, la situation est ici totalement sans espoir, car ni v_{1i} ni a fortiori w_{1i} ne suivent des lois normales, même lorsque l'échantillon est de grande taille. On ne peut dans ces conditions estimer l'équation (4) au moyen d'un modèle probit, car l'hypothèse de normalité est ici indispensable, contrairement au cas linéaire. Le recours à une estimation directe du modèle bivarié est ici incontournable.

4.2 La vraisemblance du modèle bivarié

La contribution à la vraisemblance d'un individu i correspond à la probabilité que les deux variables latentes appartiennent aux intervalles correspondant à l'observation des états :

$$L_i = \iint_{Y_{1i}^* \in D(Y_{1i}); Y_{2i}^* \in D(Y_{2i})} \exp\left(-\frac{U_{1i}^2 + U_{2i}^2 - 2\mathbf{r}U_{1i}U_{2i}}{2(1-\mathbf{r}^2)}\right) dY_{1i}^* dY_{2i}^*$$

avec
$$U_{1i} = \frac{(Y_{1i}^* - X_{1i}\mathbf{b}_1 - \mathbf{a}d_i)}{\mathbf{s}_1},$$

$$U_{2i} = \frac{(Y_{2i}^* - X_{2i}\mathbf{b}_2)}{\mathbf{s}_2},$$

$D(Y_{1i})$ (resp. $D(Y_{2i})$) représentant le domaine de variation de Y_{1i}^* (resp. Y_{2i}^*) correspondant à l'observation. Par exemple, si $Y_{1i} = 1$, alors $Y_{1i}^* \geq 0_i$ (et $U_{1i} \leq X_{1i}\mathbf{b}_1 + \mathbf{a}d_i$).

Si le logiciel dont on dispose fournit un algorithme de maximisation et la fonction cumulative de la loi normale bivariée (Gauss, SAS,...), il est relativement aisé de déterminer l'estimateur du maximum de vraisemblance correspondant au modèle. Dans le cas contraire, on peut avoir recours à la procédure d'estimation simplifiée décrite ci-dessous.

4.3 Une procédure d'estimation simplifiée

Pour estimer le modèle probit bivarié, on peut en effet reprendre une approche introduite notamment par Mroz ([4]), consistant à redéfinir les termes d'erreur dans le modèle initial :

$$u_1 = \mathbf{r}_1 v + \mathbf{e}_1$$

$$u_2 = \mathbf{r}_2 v + \mathbf{e}_2$$

où \mathbf{e}_1 , \mathbf{e}_2 et v sont supposés de moyenne nulle, indépendants entre eux et indépendants des variables explicatives du modèle. On peut interpréter cette formulation en considérant v comme une variable aléatoire inobservée qui influence linéairement les deux perturbations. Dans l'étude, on supposera que les trois variables sont distribuées selon une loi normale et que v a une variance unitaire.

Conditionnellement à la valeur prise par v , la distribution jointe de u_1 et u_2 est donnée par la formule :

$$f(u_1, u_2 / v) = \frac{1}{s_1} \mathbf{j} \left(\frac{u_1 - \mathbf{r}_1 v}{s_1} \right) \frac{1}{s_2} \mathbf{j} \left(\frac{u_2 - \mathbf{r}_2 v}{s_2} \right)$$

où s_1 et s_2 représentent les écart-types de \mathbf{e}_1 et \mathbf{e}_2 . La distribution jointe et non conditionnelle de u_1 et u_2 peut alors s'écrire :

$$f(u_1, u_2) = \int f(u_1, u_2 / v) d\Phi(v)$$

Elle correspond à la densité d'une loi normale bidimensionnelle, de moyenne nulle, de variances $(s_1^2 + \mathbf{r}_1^2)$ et $(s_2^2 + \mathbf{r}_2^2)$ et de covariance $\mathbf{r}_1 \mathbf{r}_2$. Le modèle ainsi redéfini n'est pas identifiable, mais il se présente sous la forme standard d'une intégration gaussienne. Or, ce type d'intégrale peut être

approximée avec une très grande précision au moyen d'une somme en ayant recours à la formule d'intégration d'Hermite (Butler et Moffit (1982)) :

$$\int_{-\infty}^{+\infty} g(\mathbf{n}) \exp(-v^2) d\mathbf{n} = \sum_{g=1}^G w_g g(v_g)$$

où G est le nombre de points v_g où la fonction est évaluée et w_g les poids aux différents points.

La contribution de l'individu i à la vraisemblance s'écrit alors :

$$l_i = \log \left[\sum_{g=1}^G \sqrt{2} w_g Z_{1i} \left(\frac{X_{1i} \mathbf{b}_1 + \mathbf{a} d_i + \mathbf{r}_1 \sqrt{2} v_g}{s_1} \right) Z_{2i} \left(\frac{X_{2i} \mathbf{b}_{21} + \mathbf{r}_2 \sqrt{2} v_g}{s_2} \right) \right],$$

où $Z_l(x) = \Phi(x)$ si $Y_l = 1$ et $Z_l(x) = 1 - \Phi(x)$ si $Y_l = 0$ ($l=1,2$). Le terme $\sqrt{2}$ apparaît dans l'expression du fait des différences entre la forme de la densité et la formule d'intégration précédente. Pour assurer l'identification, on peut introduire les contraintes :

$$s_2 = \mathbf{r}_2 = \frac{1}{\sqrt{2}}$$

et

$$s_1 = 1.$$

La première restriction permet d'assurer que $\mathbf{s}_2 = 1$ comme dans le modèle initial. Pour des raisons de simplification des calculs, il est plus difficile de respecter la même analogie pour \mathbf{s}_1 . Sous cette forme, le calcul de la log-vraisemblance, voire de ses dérivées est cependant relativement aisé, le recours au conditionnement permettant d'éviter le calcul d'une intégrale double.

Cette procédure peut être mise en œuvre si le logiciel dont on dispose ne fournit pas la fonction cumulative de la loi normale bivariée. Son principe d'utilisation est en outre un peu plus général, car il s'applique également au cas où la loi de v n'est pas normale. On peut notamment avoir recours à des lois non paramétriques, ceci permet notamment de tester la pertinence de la loi normale bivariée postulée a priori.

Références

- [1] Heckman J. (1974): «Shadow prices, market prices and labor supply », *Econometrica*, vol 42, 679-693.
- [2] Heckman J. (1978): «Dummy endogenous variables in a simultaneous equation system », *Econometrica*, vol 46, 931-959
- [3] Lee L-F, Maddala G., Trost R. (1980) : « Asymptotic covariance matrices of two-stage probit and two-stage tobit methods for simultaneous equations models with selectivity », *Econometrica*, vol 48, 491-503.
- [4] Mroz T. (1999): « Discrete factor approximation in simultaneous equation models : estimating the impact of a dummy endogenous variable on continuous outcome », *Journal of Econometrics*, vol 92, 233-274.
- [5] Pagan A. (1984) : “Econometric issues in the generating of regressions with generated regressors”, *International Economic Review*, 25, n°1, pp 221-247.
- [6] Robin J.M. (2000) : « Modèles structurels et variables explicatives endogènes »; *Méthodologie statistique*, n°2002, INSEE.